

*Bernoulli* **18**(3), 2012, 1002–1030  
DOI: [10.3150/11-BEJ356](https://doi.org/10.3150/11-BEJ356)

# A class of measure-valued Markov chains and Bayesian nonparametrics

STEFANO FAVARO<sup>1</sup>, ALESSANDRA GUGLIELMI<sup>2</sup> and STEPHEN G. WALKER<sup>3</sup>

<sup>1</sup>*Università di Torino and Collegio Carlo Alberto, Dipartimento di Statistica e Matematica Applicata, Corso Unione Sovietica 218/bis, 10134 Torino, Italy. E-mail: [favaro@econ.unito.it](mailto:favaro@econ.unito.it)*

<sup>2</sup>*Politecnico di Milano, Dipartimento di Matematica, P.zza Leonardo da Vinci 32, 20133 Milano, Italy. E-mail: [alessandra.guglielmi@polimi.it](mailto:alessandra.guglielmi@polimi.it)*

<sup>3</sup>*University of Kent, Institute of Mathematics, Statistics and Actuarial Science, Canterbury CT27NZ, UK. E-mail: [S.G.Walker@kent.ac.uk](mailto:S.G.Walker@kent.ac.uk)*

Measure-valued Markov chains have raised interest in Bayesian nonparametrics since the seminal paper by (*Math. Proc. Cambridge Philos. Soc.* **105** (1989) 579–585) where a Markov chain having the law of the Dirichlet process as unique invariant measure has been introduced. In the present paper, we propose and investigate a new class of measure-valued Markov chains defined via exchangeable sequences of random variables. Asymptotic properties for this new class are derived and applications related to Bayesian nonparametric mixture modeling, and to a generalization of the Markov chain proposed by (*Math. Proc. Cambridge Philos. Soc.* **105** (1989) 579–585), are discussed. These results and their applications highlight once again the interplay between Bayesian nonparametrics and the theory of measure-valued Markov chains.

*Keywords:* Bayesian nonparametrics; Dirichlet process; exchangeable sequences; linear functionals of Dirichlet processes; measure-valued Markov chains; mixture modeling; Pólya urn scheme; random probability measures

## 1. Introduction

Measure-valued Markov chains, or more generally measure-valued Markov processes, arise naturally in modeling the composition of evolving populations and play an important role in a variety of research areas such as population genetics and bioinformatics (see, e.g., [5, 9, 10, 26]), Bayesian nonparametrics [31, 38], combinatorics [26] and statistical physics [5, 6, 26]. In particular, in Bayesian nonparametrics there has been interest in measure-valued Markov chains since the seminal paper by [12], where the law of the Dirichlet process has been characterized as the unique invariant measure of a certain measure-valued Markov chain.

In order to introduce the result by [12], let us consider a Polish space  $\mathbb{X}$  endowed with the Borel  $\sigma$ -field  $\mathcal{X}$  and let  $\mathcal{P}_{\mathbb{X}}$  be the space of probability measures on  $\mathbb{X}$  with the  $\sigma$ -

This is an electronic reprint of the original article published by the ISI/BS in *Bernoulli*, 2012, Vol. 18, No. 3, 1002–1030. This reprint differs from the original in pagination and typographic detail.

field  $\mathcal{P}_{\mathbb{X}}$  generated by the topology of weak convergence. If  $\alpha$  is a strictly positive finite measure on  $\mathbb{X}$  with total mass  $a > 0$ ,  $Y$  is a  $\mathbb{X}$ -valued random variable (r.v.) distributed according to  $\alpha_0 := \alpha/a$  and  $\theta$  is a r.v. independent of  $Y$  and distributed according to a Beta distribution with parameter  $(1, a)$  then, Theorem 3.4 in [33] implies that a Dirichlet process  $P$  on  $\mathbb{X}$  with parameter  $\alpha$  uniquely satisfies the distributional equation

$$P \stackrel{d}{=} \theta \delta_Y + (1 - \theta)P, \quad (1)$$

where all the random elements on the right-hand side of (1) are independent. All the r.v.s introduced in this paper are meant to be assigned on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  unless otherwise stated. In [12], (1) is recognized as the distributional equation for the unique invariant measure of a measure-valued Markov chain  $\{P_m, m \geq 0\}$  defined via the recursive identity

$$P_m = \theta_m \delta_{Y_m} + (1 - \theta_m)P_{m-1}, \quad m \geq 1, \quad (2)$$

where  $P_0 \in \mathcal{P}_{\mathbb{X}}$  is arbitrary,  $\{Y_m, m \geq 1\}$  is a sequence of  $\mathbb{X}$ -valued r.v.s independent and identically distributed as  $Y$  and  $\{\theta_m, m \geq 1\}$  is a sequence of r.v.s, independent and identically distributed as  $\theta$  and independent of  $\{Y_m, m \geq 1\}$ . We term  $\{P_m, m \geq 0\}$  as the Feigin–Tweedie Markov chain. By investigating the functional Markov chain  $\{G_m, m \geq 0\}$ , with  $G_m := \int_{\mathbb{X}} g(x) P_m(dx)$  for any  $m \geq 0$  and for any measurable linear function  $g: \mathbb{X} \mapsto \mathbb{R}$ , [12] provide properties of the corresponding linear functional of a Dirichlet process. In particular, the existence of the linear functional  $G := \int_{\mathbb{X}} g(x) P(dx)$  of the Dirichlet process  $P$  is characterized according to the condition  $\int_{\mathbb{X}} \log(1 + |g(x)|) \alpha(dx) < +\infty$ ; these functionals were considered by [16] and their existence was also investigated by [7] who referred to them as moments, as well as by [39] and [4]. Further developments of the linear functional Markov chain  $\{G_m, m \geq 0\}$  are provided by [15, 17] and more recently by [8].

Starting from the distributional equation (1), a constructive definition of the Dirichlet process has been proposed by [33]. If  $P$  is a Dirichlet process on  $\mathbb{X}$  with parameter  $\alpha = a\alpha_0$ , then  $P = \sum_{1 \leq i \leq \infty} p_i \delta_{Y_i}$  where  $\{Y_i, i \geq 1\}$  is a sequence of independent r.v.s identically distributed according to  $\alpha_0$  and  $\{p_i, i \geq 1\}$  is a sequence of r.v.s independent of  $\{Y_i, i \geq 1\}$  and derived by the so-called stick breaking construction, that is,  $p_1 = w_1$  and  $p_i = w_i \prod_{1 \leq j \leq i-1} (1 - w_j)$  for  $i > 1$ , with  $\{w_i, i \geq 1\}$  being a sequence of independent r.v.s identically distributed according to a Beta distribution with parameter  $(1, a)$ . Then, equation (1) arises by considering

$$P = p_1 \delta_{Y_1} + (1 - w_1) \sum_{i=2}^{\infty} \tilde{p}_i \delta_{Y_i},$$

where now  $\tilde{p}_2 = w_2$  and  $\tilde{p}_i = w_i \prod_{2 \leq j \leq i-1} (1 - w_j)$  for  $i > 2$ . Thus, it is easy to see that  $\tilde{P} := \sum_{2 \leq i \leq \infty} \tilde{p}_i \delta_{Y_i}$  is also a Dirichlet process on  $\mathbb{X}$  with parameter  $\alpha$  and it is independent of the pairs of r.v.s  $(p_1, Y_1)$ . If we would extend this idea to  $n$  initial samples, we should

consider writing

$$P = \theta \sum_{i=1}^n \left( \frac{p_i}{\theta} \right) \delta_{Y_i} + (1 - \theta) \tilde{P},$$

where  $\theta = \sum_{1 \leq i \leq n} p_i = 1 - \prod_{1 \leq i \leq n} (1 - w_i)$  and  $\tilde{P}$  is a Dirichlet process on  $\mathbb{X}$  with parameter  $\alpha$  independent of the random vectors  $(p_1, \dots, p_n)$  and  $(Y_1, \dots, Y_n)$ . However, this is not an easy extension since the distribution of  $\theta$  is unclear, and moreover  $\theta$  and  $\sum_{1 \leq i \leq n} (p_i/\theta) \delta_{Y_i}$  are not independent. For this reason, in [11] an alternative distributional equation has been introduced. Let  $\alpha$  be a strictly positive finite measure on  $\mathbb{X}$  with total mass  $a > 0$  and let  $\{Y_j, j \geq 1\}$  be a  $\mathbb{X}$ -valued Pólya sequence with parameter  $\alpha$  (see [2]), that is,  $\{Y_j, j \geq 1\}$  is a sequence of  $\mathbb{X}$ -valued r.v.s characterized by the following predictive distributions

$$\mathbb{P}(Y_{j+1} \in A | Y_1, \dots, Y_j) = \frac{1}{a+j} \alpha(A) + \frac{1}{a+j} \sum_{i=1}^j \delta_{Y_i}(A), \quad j \geq 1,$$

and  $\mathbb{P}(Y_1 \in A) = \alpha(A)/a$ , for any  $A \in \mathcal{X}$ . The sequence  $\{Y_j, j \geq 1\}$  is exchangeable, that is, for any  $j \geq 1$  and any permutation  $\sigma$  of the indexes  $(1, \dots, j)$ , the law of the r.v.s  $(Y_1, \dots, Y_j)$  and  $(Y_{\sigma(1)}, \dots, Y_{\sigma(j)})$  coincide; in particular, according to the celebrated de Finetti representation theorem, the Pólya sequence  $\{Y_j, j \geq 1\}$  is characterized by a so-called de Finetti measure, which is the law of a Dirichlet process on  $\mathbb{X}$  with parameter  $\alpha$ . For a fixed integer  $n \geq 1$ , let  $(q_1^{(n)}, \dots, q_n^{(n)})$  be a random vector distributed according to the Dirichlet distribution with parameter  $(1, \dots, 1)$ ,  $\sum_{1 \leq i \leq n} q_i^{(n)} = 1$ , and let  $\theta$  be a r.v. distributed according to a Beta distribution with parameter  $(n, a)$  such that  $\{Y_i, i \geq 1\}$ ,  $(q_1^{(n)}, \dots, q_n^{(n)})$  and  $\theta$  are mutually independent. Moving from such a collection of random elements, Lemma 1 in [11] implies that a Dirichlet process  $P^{(n)}$  on  $\mathbb{X}$  with parameter  $\alpha$  uniquely satisfy the distributional equation

$$P^{(n)} \stackrel{d}{=} \theta \sum_{i=1}^n q_i^{(n)} \delta_{Y_i} + (1 - \theta) P^{(n)}, \quad (3)$$

where all the random elements on the right-hand side of (3) are independent. In order to emphasize the additional parameter  $n$ , we used an upper-script  $(n)$  on the Dirichlet process  $P$  and on the random vector  $(q_1^{(n)}, \dots, q_n^{(n)})$ . It can be easily checked that equation (3) generalizes (1), which can be recovered by setting  $n = 1$ .

In the present paper, our aim is to further investigate the distributional equation (3) and its implications in Bayesian nonparametrics theory and methods. The first part of the paper is devoted to investigate the random element  $\sum_{1 \leq i \leq n} q_i^{(n)} \delta_{Y_i}$  in (3) which is recognized to be the random probability measure (r.p.m.) at the  $n$ th step of a measure-valued Markov chain defined via the recursive identity

$$\sum_{i=1}^n q_i^{(n)} \delta_{Y_i} = W_n \delta_{Y_n} + (1 - W_n) \sum_{i=1}^{n-1} q_i^{(n-1)} \delta_{Y_i}, \quad n \geq 1, \quad (4)$$

where  $\{W_n, n \geq 1\}$  is a sequence of independent r.v.s, each  $W_n$  distributed according to a Beta distribution with parameter  $(1, n-1)$ ,  $q_i^{(n)} = W_i \prod_{i+1 \leq j \leq n} (1 - W_j)$  for  $i = 1, \dots, n$  and  $n \geq 1$  and the sequence  $\{W_n, n \geq 1\}$  is independent from  $\{Y_n, n \geq 1\}$ . More generally, we observe that the measure-valued Markov chain defined via the recursive identity (4) can be extended by considering, instead of a Pólya sequence  $\{Y_n, n \geq 1\}$  with parameter  $\alpha$ , any exchangeable sequence  $\{Z_n, n \geq 1\}$  characterized by some de Finetti measure on  $\mathcal{P}_{\mathbb{X}}$  and such that  $\{W_n, n \geq 1\}$  is independent from  $\{Z_n, n \geq 1\}$ . Asymptotic properties for this new class of measure-valued Markov chains are derived and some linkages to Bayesian nonparametric mixture modelling are discussed. In particular, we remark how it is closely related to a well-known recursive algorithm introduced in [25] for estimating the underlying mixing distribution in mixture models, the so-called Newton's algorithm.

In the second part of the paper, by using finite and asymptotic properties of the r.p.m.  $\sum_{1 \leq i \leq n} q_i^{(n)} \delta_{Y_i}$  and by following the original idea of [12], we define and investigate from (3) a class of measure-valued Markov chain  $\{P_m^{(n)}, m \geq 0\}$  which generalizes the Feigin–Tweedie Markov chain, introducing a fixed integer parameter  $n$ . Our aim is in providing features of the Markov chain  $\{P_m^{(n)}, m \geq 0\}$  in order to verify if it preserves some of the properties characterizing the Feigin–Tweedie Markov chain; furthermore, we are interested in analyzing asymptotic (as  $m$  goes to  $+\infty$ ) properties of the associated linear functional Markov chain  $\{G_m^{(n)}, m \geq 0\}$  with  $G_m^{(n)} := \int_{\mathbb{X}} g(x) P_m^{(n)}(dx)$  for any  $m \geq 0$  and for any function  $g: \mathbb{X} \mapsto \mathbb{R}$  such that  $\int_{\mathbb{X}} \log(1 + |g(x)|) \alpha(dx) < +\infty$ . In particular, we show that the Feigin–Tweedie Markov chain  $\{P_m, m \geq 0\}$  sits in a larger class of measure-valued Markov chains  $\{P_m^{(n)}, m \geq 0\}$  parametrized by an integer number  $n$  and still having the law of a Dirichlet process with parameter  $\alpha$  as unique invariant measure. The role of the further parameter  $n$  is discussed in terms of new potential applications of the Markov chain  $\{P_m^{(n)}, m \geq 0\}$  with respect to the the known applications of the Feigin–Tweedie Markov chain.

Following these guidelines, in Section 2 we introduce a new class of measure-valued Markov chains  $\{Q_n, n \geq 1\}$  defined via exchangeable sequences of r.v.s; asymptotic results for  $\{Q_n, n \geq 1\}$  are derived and applications related to Bayesian nonparametric mixture modelling are discussed. In Section 3, we show that the Feigin–Tweedie Markov chain  $\{P_m, m \geq 0\}$  sits in a larger class of measure-valued Markov chains  $\{P_m^{(n)}, m \geq 0\}$ , which is investigated in comparison with  $\{P_m, m \geq 0\}$ . In Section 4, some concluding remarks and future research lines are presented.

## 2. A class of measure-valued Markov chains and Newton's algorithm

Let  $\{W_n, n \geq 1\}$  be a sequence of independent r.v.s such that  $W_1 = 1$  almost surely and  $W_n$  has Beta distribution with parameter  $(1, n-1)$  for  $n \geq 2$ . Moreover, let  $\{Z_n, n \geq 1\}$  be a sequence of  $\mathbb{X}$ -valued exchangeable r.v.s independent from  $\{W_n, n \geq 1\}$  and characterized by some de Finetti measure on  $\mathcal{P}_{\mathbb{X}}$ . Let us consider the measure-valued

Markov chain  $\{Q_n, n \geq 1\}$  defined via the recursive identity

$$Q_n = W_n \delta_{Z_n} + (1 - W_n) Q_{n-1}, \quad n \geq 1. \quad (5)$$

In the next theorem, we provide an alternative representation of  $Q_n$  and show that  $Q_n(\omega)$  converges weakly to some limit probability  $Q(\omega)$  for almost all  $\omega \in \Omega$ , *that is*, for each  $\omega$  in some set  $A \in \mathcal{F}$  with  $\mathbb{P}(A) = 1$ . In short, we use notation  $Q_n \Rightarrow Q$  a.s.- $\mathbb{P}$ .

**Theorem 1.** *Let  $\{Q_n, n \geq 1\}$  be the Markov chain defined by (5). Then:*

- (i) *an equivalent representation of  $Q_n$ ,  $n = 1, 2, \dots$  is*

$$Q_n = \sum_{i=1}^n q_i^{(n)} \delta_{Z_i}, \quad (6)$$

*where  $\sum_{1 \leq i \leq n} q_i^{(n)} = 1$ ,  $q^{(n)} = (q_1^{(n)}, \dots, q_n^{(n)})$  has Dirichlet distribution with parameter  $(1, 1, \dots, 1)$ , and  $\{q^{(n)}, n \geq 1\}$  and  $\{Z_n, n \geq 1\}$  are independent.*

- (ii) *There exists a r.p.m.  $Q$  on  $(\Omega, \mathcal{F}, \mathbb{P})$  such that, as  $n \rightarrow +\infty$ ,*

$$Q_n \Rightarrow Q, \quad \text{a.s.-}\mathbb{P},$$

*where the law of  $Q$  is the de Finetti measure of the sequence  $\{Z_n, n \geq 1\}$ .*

**Proof.** As far as (i) is concerned, by repeated application of the recursive identity (5), it can be checked that, for any  $n \geq 1$ ,

$$Q_n = \sum_{i=1}^n W_i \prod_{j=i+1}^n (1 - W_j) \delta_{Z_i},$$

where  $W_1 = 1$  almost surely and  $\prod_{i+1 \leq j \leq n} (1 - W_j)$  is defined to be 1 when  $i = n$ . Defining  $q_i^{(n)} := W_i \prod_{i+1 \leq j \leq n} (1 - W_j)$ ,  $i = 1, \dots, n$ , it is straightforward to show that  $q^{(n)} = (q_1^{(n)}, \dots, q_n^{(n)})$  has the Dirichlet distribution with parameter  $(1, 1, \dots, 1)$  and  $\sum_{1 \leq i \leq n} q_i^{(n)} = 1$ , so that (6) holds.

Regarding (ii), by the definition of the Dirichlet distribution, an equivalent representation of (6) is

$$Q_n = \sum_{i=1}^n q_i^{(n)} \delta_{Z_i} = \sum_{i=1}^n \frac{\lambda_i}{\sum_{j=1}^n \lambda_j} \delta_{Z_i},$$

where  $\{\lambda_n, n \geq 1\}$  is a sequence of r.v.s independent and identically distributed according to standard exponential distribution, independent from  $\{Z_n, n \geq 1\}$ . Let  $g: \mathbb{X} \rightarrow \mathbb{R}$  be any bounded continuous function, and consider

$$G_n = \int_{\mathbb{X}} g dQ_n = \sum_{i=1}^n \frac{\lambda_i}{\sum_{j=1}^n \lambda_j} g(Z_i) = \frac{\sum_{i=1}^n \lambda_i g(Z_i) / n}{\sum_{i=1}^n \lambda_i / n}.$$

The expression in the denominator converges almost surely to 1 by the strong law of large numbers. As far as the numerator is concerned, let  $Q$  be the r.p.m. defined on  $(\Omega, \mathcal{F}, \mathbb{P})$ , such that the r.v.s  $\{Z_n, n \geq 1\}$  are independent and identically distributed conditionally on  $Q$ ; the existence of such a random element is guaranteed by the de Finetti representation theorem (see, e.g., [32], Theorem 1.49). It can be shown that  $\{\lambda_n g(Z_n), n \geq 1\}$  is a sequence of exchangeable r.v.s and, if  $t_1, \dots, t_n \in \mathbb{R}$ ,

$$\begin{aligned} & \mathbb{P}(\lambda_1 g(Z_1) \leq t_1, \dots, \lambda_n g(Z_n) \leq t_n) \\ &= \int_{(0, +\infty)^n} \mathbb{P}(\lambda_1 g(Z_1) \leq t_1, \dots, \lambda_n g(Z_n) \leq t_n | \lambda_1, \dots, \lambda_n) \prod_{i=1}^n e^{-\lambda_i} d\lambda_i \\ &= \int_{(0, +\infty)^n} \int_{\mathcal{P}_{\mathbb{R}}} \prod_{i=1}^n F_{Q^*} \left( \frac{t_i}{\lambda_i} \right) \mu(dQ^*) \prod_{i=1}^n e^{-\lambda_i} d\lambda_i \\ &= \int_{\mathcal{P}_{\mathbb{R}}} \prod_{i=1}^n \left( \int_0^{+\infty} F_{Q^*} \left( \frac{t_i}{\lambda_i} \right) e^{-\lambda_i} d\lambda_i \right) \mu(dQ^*), \end{aligned}$$

where  $Q^*(A, \omega) := P(g^{-1}(A), \omega)$ ,  $\omega \in \Omega$ ,  $A \in \mathcal{R}$ , is a r.p.m. with trajectories in  $\mathcal{P}_{\mathbb{R}}$ , and  $F_{Q^*}$  denotes the random distribution relative to  $Q^*$ . This means that, conditionally on  $Q^*$ ,  $\{\lambda_n g(Z_n), n \geq 1\}$  is a sequence of r.v.s independent and identically distributed according to the random distribution (evaluated in  $t$ )

$$\int_0^{+\infty} F_{Q^*} \left( \frac{t}{y} \right) e^{-y} dy.$$

Of course,  $\mathbb{E}|\lambda_1 g(Z_1)| = \mathbb{E}(\lambda_1) \mathbb{E}|g(Z_1)| < +\infty$  since  $g$  is bounded. As in [3], Example 7.3.1, this condition implies

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \lambda_i g(Z_i) \xrightarrow{\text{a.s.}} \mathbb{E}(\lambda_1 g(Z_1) | Q^*) = \int_{\mathbb{R}} t d \left( \int_0^{+\infty} F_{Q^*} \left( \frac{t}{y} \right) e^{-y} dy \right) \\ &= \int_{\mathbb{R}} u Q^*(du) = \int_{\mathbb{X}} g(x) Q(dx), \end{aligned}$$

so that  $G_n \rightarrow \int_{\mathbb{X}} g(x) Q(dx)$  a.s.- $\mathbb{P}$ . By Theorem 2.2 in [1], it follows that  $Q_n \Rightarrow Q$  a.s.- $\mathbb{P}$  as  $n \rightarrow +\infty$ .  $\square$

Throughout the paper,  $\alpha$  denotes a strictly positive and finite measure on  $\mathbb{X}$  with total mass  $a$ , unless otherwise stated. If the exchangeable sequence  $\{Z_n, n \geq 1\}$  is the Pólya sequence with parameter  $\alpha$ , then by Theorem 1(i)  $\{Q_n, n \geq 1\}$  is the Markov chain defined via the recursive identity (4); in particular, by Theorem 1(ii),  $Q_n \Rightarrow Q$  a.s.- $\mathbb{P}$  where  $Q$  is a Dirichlet process on  $\mathbb{X}$  with parameter  $\alpha$ . This means that, for any fixed integer  $n \geq 1$ , the r.p.m.  $Q_n$  can be interpreted as an approximation of a Dirichlet process with parameter  $\alpha$ . In Appendix A.1, we present an alternative proof of the weak convergence

(convergence of the finite dimensional distribution) of  $\{Q_n, n \geq 1\}$  to a Dirichlet process on  $\mathbb{X}$  with parameter  $\alpha$ , using a combinatorial technique. As a byproduct of this proof, we obtain an explicit expression for the moment of order  $(r_1, \dots, r_k)$  of the  $k$ -dimensional Pólya distribution.

A straightforward generalization of the Markov chain  $\{Q_n, n \geq 0\}$  can be obtained by considering a nonparametric hierarchical mixture model. Let  $k: \mathbb{X} \times \Theta \rightarrow \mathbb{R}^+$  be a kernel, that is,  $k(x, \vartheta)$  is a measurable function such that  $x \mapsto k(x, \vartheta)$  is a density with respect to some  $\sigma$ -finite measure  $\lambda$  on  $\mathbb{X}$ , for any fixed  $\vartheta \in \Theta$ , where  $\Theta$  is a Polish space (with the usual Borel  $\sigma$ -field). Let  $\{Q_n, n \geq 1\}$  be the Markov chain defined via (5). Then for each  $x \in \mathbb{X}$  we introduce a real-valued Markov chain  $\{f_n^{(Q)}(x), n \geq 1\}$  defined via the recursive identity

$$f_n^{(Q)}(x) = W_n k(x, \vartheta_n) + (1 - W_n) f_{n-1}^{(Q)}(x), \quad n \geq 1, \quad (7)$$

where

$$f_n^{(Q)}(x) = \int_{\Theta} k(x, \vartheta) Q_n(d\vartheta).$$

By a straightforward application of Theorem 2.2 in [1], for any fixed  $x \in \mathbb{X}$ , when  $\vartheta \mapsto k(x, \vartheta)$  is continuous for all  $x \in \mathbb{X}$  and bounded by a function  $h(x)$ , as  $n \rightarrow +\infty$ , then

$$f_n^{(Q)}(x) \rightarrow f^{(Q)}(x) := \int_{\Theta} k(x, \vartheta) Q(d\vartheta), \quad \text{a.s.-}\mathbb{P}, \quad (8)$$

where  $Q$  is the limit r.p.m. in Theorem 1. For instance, if  $Q$  is a Dirichlet process on  $\mathbb{X}$  with parameter  $\alpha$ ,  $f^{(Q)}$  is precisely the density in the Dirichlet process mixture model introduced by [19]. When  $h(x)$  is a  $\lambda$ -integrable function, not only the limit  $f^{(Q)}(x)$  is a random density, but a stronger result than (8) is achieved.

**Theorem 2.** *If  $\vartheta \mapsto k(x, \vartheta)$  is continuous for all  $x \in \mathbb{X}$  and bounded by a  $\lambda$ -integrable function  $h(x)$ , then*

$$\lim_{n \rightarrow +\infty} \int_{\mathbb{X}} |f_n^{(Q)}(x) - f^{(Q)}(x)| \lambda(dx) \rightarrow 0, \quad \text{a.s.-}\mathbb{P},$$

where  $Q$  is the limit r.p.m. in Theorem 1.

**Proof.** The functions  $f_n^{(Q)}$  and  $f^{(Q)}(x) = \int_{\Theta} k(x, \vartheta) Q(d\vartheta)$ , defined on  $\mathbb{X} \times \Omega$ , are  $\mathcal{X} \otimes \mathcal{F}$ -measurable, by a monotone class argument. In fact, by kernel's definition,  $(x, \vartheta) \mapsto k(x, \vartheta)$  is  $\mathcal{X} \otimes \mathcal{B}(\Theta)$ -measurable. Moreover, if  $k = 1_A 1_B$ ,  $A \in \mathcal{X}$  and  $B \in \mathcal{B}(\Theta)$ , then

$$f^{(Q)}(x, \omega) = \int k(x, \vartheta) Q(d\vartheta; \omega) = 1_A(x) Q(B; \omega)$$

is  $\mathcal{X} \otimes \mathcal{F}$ -measurable. Let  $\mathcal{C} = \{C \in \mathcal{X} \otimes \mathcal{B}(\Theta): \int 1_C(x, \vartheta) Q(d\vartheta; \omega) \text{ is } \mathcal{X} \otimes \mathcal{F}\text{-measurable}\}$ . Since  $\mathcal{C}$  contains the rectangles, it contains the field generated by rectangles, and, since  $\mathcal{C}$

is a monotone class,  $\mathcal{C} = \mathcal{X} \otimes \mathcal{B}(\Theta)$ . The assertion holds for  $f^{(Q)}$  of the form

$$f^{(Q)}(x) = \int_{\Theta} k(x, \vartheta) Q(d\vartheta)$$

since there exist a sequence of simple function on rectangles which converges pointwise to  $k$ . Therefore,  $A := \{(\omega, x): f_n^{(Q)}(\omega, x) \text{ does not converge to } f^{(Q)}(\omega, x)\} \in \mathcal{F} \otimes \mathcal{X}$ . Then, by Fubini's theorem,

$$\begin{aligned} & \int \lambda\{x: f_n^{(Q)}(\omega, x) \text{ does not converge to } f^{(Q)}(\omega, x)\} \mathbb{P}(d\omega) \\ &= \int \int 1_A(\omega, x) \lambda(dx) \mathbb{P}(d\omega) \\ &= \int \mathbb{P}\{\omega: f_n^{(Q)}(\omega, x) \text{ does not converge to } f^{(Q)}(\omega, x)\} \lambda(dx) \\ &= \int 0 \lambda(dx) = 0. \end{aligned}$$

Hence,  $\mathbb{P}(H) = 1$  where  $H$  is the set of  $\omega$  such that  $\lambda\{x: f_n^{(Q)}(\omega, x) \text{ does not converge to } f^{(Q)}(\omega, x)\} = 0$ . For any  $\omega$  fixed in  $H$ , it holds  $f_n^{(Q)}(\omega, \cdot) \rightarrow f^{(Q)}(\omega, \cdot)$ ,  $\lambda$ -a.e., so that by the Scheffé's theorem we have

$$\lim_{n \rightarrow +\infty} \int_{\mathbb{X}} |f_n^{(Q)}(\omega, x) - f^{(Q)}(\omega, x)| \lambda(dx) \rightarrow 0.$$

The theorem follows since  $\mathbb{P}(H) = 1$ .  $\square$

We conclude this section by remarking an interesting linkage between the Markov chain  $\{Q_n, n \geq 1\}$  and the so-called Newton's algorithm, originally introduced in [25] for estimating the mixing density when a finite sample is available from the corresponding mixture model. See also [24] and [23]. Briefly, suppose that  $X_1, \dots, X_n$  are  $n$  r.v.s independent and identically distributed according to the density function

$$\tilde{f}(x) = \int_{\Theta} k(x, \vartheta) \tilde{Q}(d\vartheta), \quad (9)$$

where  $k(x, \vartheta)$  is a known kernel dominated by a  $\sigma$ -finite measure  $\lambda$  on  $\mathbb{X}$ ; assume that the mixing distribution  $\tilde{Q}$  is absolutely continuous with respect to some  $\sigma$ -finite measure  $\mu$  on  $\Theta$ . [23] proposed to estimate  $\tilde{q} = d\tilde{Q}/d\mu$  as follows: fix an initial estimate  $\hat{q}_1$  and a sequence of weights  $w_1, w_2, \dots, w_n \in (0, 1)$ . Given  $X_1, \dots, X_n$  independent and identically distributed observations from  $\tilde{f}$ , compute

$$\hat{q}_i(\vartheta) = (1 - w_i) \hat{q}_{i-1}(\vartheta) + w_i \frac{k(x_i, \vartheta) \hat{q}_{i-1}(\vartheta)}{\int_{\Theta} k(x_i, \vartheta) \hat{q}_{i-1}(\vartheta) \mu(d\vartheta)}, \quad \vartheta \in \Theta$$



for  $i = 2, 3, \dots, n$  and produce  $\hat{q}_n$  as the final estimate. We refer to [13, 20, 34], and [21] for a recent wider investigation of the Newton's algorithm. Here we show how the Newton's algorithm is connected to the measure-valued Markov chain  $\{Q_n, n \geq 1\}$ .

Let us consider  $n$  observations from the nonparametric hierarchical mixture model, that is,  $X_i|\vartheta_i \sim k(\cdot, \vartheta_i)$  and  $\vartheta_i|Q \sim Q$  where  $Q$  is a r.p.m. If we observed  $\{\vartheta_i, i \geq 1\}$ , then by virtue of (ii) in Theorem 1, we could construct a sequence of distributions

$$Q_i = W_i \delta_{\vartheta_i} + (1 - W_i) Q_{i-1}, \quad i = 1, \dots, n$$

for estimating the limit r.p.m.  $Q$ , where  $\{W_i, i \geq 1\}$  is a sequence of independent r.v.s such that  $W_1 = 1$  almost surely and  $W_i$  has Beta distribution with parameters  $(1, i - 1)$ . This approximating sequence is precisely the sequence (5). Therefore, taking the expectation of both sides of the previous recursive equation, and defining  $\tilde{Q}_i := \mathbb{E}[Q_i]$ ,  $w_i = \mathbb{E}[W_i] = 1/i$ , we have

$$\tilde{Q}_i = w_i \delta_{\vartheta_i} + (1 - w_i) \tilde{Q}_{i-1}, \quad i = 1, \dots, n, \quad (10)$$

which can represent a predictive distribution for  $\vartheta_{i+1}$ , and hence an estimate for  $Q$ .

However, instead of observing the sequence  $\{\vartheta_i, i \geq 1\}$ , it is actually the sequence  $\{X_i, i \geq 1\}$  which is observed; in particular, we can assume that  $X_1, \dots, X_n$  are  $n$  r.v.s independent and identically distributed according to the density function (9). Therefore, instead of (10), we consider

$$\tilde{Q}_i(d\vartheta) = (1 - w_i) \tilde{Q}_{i-1}(d\vartheta) + w_i \frac{k(x_i, \vartheta) \tilde{Q}_{i-1}(d\vartheta)}{\int_{\Theta} k(x_i, \vartheta) \tilde{Q}_{i-1}(d\vartheta)}, \quad i = 1, \dots, n,$$

where  $\delta_{\vartheta_i}$  in (10) has been substituted (or estimated, if you prefer) by  $k(x_i, \vartheta) \tilde{Q}_{i-1}(d\vartheta) / \int_{\Theta} k(x_i, \vartheta) \tilde{Q}_{i-1}(d\vartheta)$ . Finally, observe that, if  $\tilde{Q}_i$  is absolutely continuous, with respect to some  $\sigma$ -finite measure  $\mu$  on  $\Theta$ , with density  $\tilde{q}_i$ , for  $i = 1, \dots, n$ , then we can write

$$\tilde{q}_i(\vartheta) = (1 - w_i) \tilde{q}_{i-1}(\vartheta) + w_i \frac{k(x_i, \vartheta) \tilde{q}_{i-1}(\vartheta)}{\int_{\Theta} k(x_i, \vartheta) \tilde{q}_{i-1}(\vartheta) \mu(d\vartheta)}, \quad i = 1, \dots, n, \quad (11)$$

which is precisely a recursive estimator of a mixing distribution proposed by [23] when the weights are fixed to be  $w_i = 1/i$  for  $i = 1, \dots, n$  and the initial estimate is  $\mathbb{E}[\delta_{\vartheta_1}]$ .

### 3. A generalized Feigin–Tweedie Markov chain

In this section our aim is to define and investigate a class of measure-valued Markov chain which generalizes the Feigin–Tweedie Markov chain introducing a fixed integer parameter  $n$ , and still has the law of a Dirichlet process with parameter  $\alpha$  as the unique invariant measure. The starting point is the distributional equation (3) introduced by [11]; see Appendix A.2 for an alternative proof of the solution of the distributional equation (3). All the proofs of Theorems in this section are in Appendix A.3 for the ease of reading.

For a fixed integer  $n \geq 1$ , let  $\theta := \{\theta_m, m \geq 1\}$  be a sequence of independent r.v.s with Beta distribution with parameter  $(n, a)$ ,  $q^{(n)} := \{(q_{m,1}^{(n)}, \dots, q_{m,n}^{(n)}), m \geq 1\}$ , with  $\sum_{1 \leq i \leq n} q_{m,i}^{(n)} = 1$  for any  $m > 0$ , be a sequence of independent r.v.s identically distributed according to a Dirichlet distribution with parameter  $(1, \dots, 1)$  and  $Y := \{(Y_{m,1}, \dots, Y_{m,n}), m \geq 1\}$  be sequence of independent r.v.s from a Pólya sequence with parameter  $\alpha$ . Moving from such collection of random elements, for each fixed integer  $n \geq 1$  we define the measure-valued Markov chain  $\{P_m^{(n)}, m \geq 0\}$  via the recursive identity

$$P_m^{(n)} = \theta_m \sum_{i=1}^n q_{m,i}^{(n)} \delta_{Y_{m,i}} + (1 - \theta_m) P_{m-1}^{(n)}, \quad m \geq 1, \quad (12)$$

where  $P_0^{(n)} \in \mathcal{P}_{\mathbb{X}}$  is arbitrary. By construction, the Markov chain  $\{P_m, m \geq 0\}$  proposed by [12] and defined via the recursive identity (2) can be recovered from  $\{P_m^{(n)}, m \geq 0\}$  by setting  $n = 1$ . Following the original idea of [12], by equation (12) we have defined the Markov chain  $\{P_m^{(n)}, m \geq 0\}$  from a distributional equation having as the unique solution the Dirichlet process. In particular, the Markov chain  $\{P_m^{(n)}, m \geq 0\}$  is defined from the distributional equation (3) which generalizes (1) substituting the random probability measure  $\delta_Y$  with the random convex linear combination  $\sum_{1 \leq i \leq n} q_i^{(n)} \delta_{Y_i}$ , for any fixed positive integer  $n$ . Observe that  $\sum_{1 \leq i \leq n} q_i^{(n)} \delta_{Y_i}$  is an example of the r.p.m.  $Q_n$  defined in (6) and investigated in the previous section, when  $\{Z_i\}$  is given by the Pólya sequence  $\{Y_i\}$  with parameter  $\alpha$ . In particular, Theorem 1 shows that  $Q_n$  a.s.-converges to the Dirichlet process  $P$  when  $n$  goes to infinity; however here we assume a different perspective, that is,  $n$  is fixed.

As for the case  $n = 1$ , the following result holds.

**Theorem 3.** *The Markov chain  $\{P_m^{(n)}, m \geq 0\}$  has a unique invariant measure  $\Pi$  which is the law of a Dirichlet process  $P$  with parameter  $\alpha$ .*

Another property which still holds in the more general case when  $n \geq 1$  is the Harris ergodicity of the functional Markov chain  $\{G_m^{(n)}, m \geq 0\}$ , under assumption (13) below. This condition is equivalent to the finiteness of the r.v.  $\int_{\mathbb{X}} |g(x)| P(dx)$ ; see also [4].

**Theorem 4.** *Let  $g: \mathbb{X} \mapsto \mathbb{R}$  be any measurable function. If*

$$\int_{\mathbb{X}} \log(1 + |g(x)|) \alpha(dx) < +\infty, \quad (13)$$

*then the Markov chain  $\{G_m^{(n)}, m \geq 0\}$  is Harris ergodic with unique invariant measure  $\Pi_g$ , which is the law of the random Dirichlet mean  $\int_{\mathbb{X}} g(x) P(dx)$ .*

We conclude the analysis of the Markov chain  $\{P_m^{(n)}, m \geq 0\}$  by providing some results on the ergodicity of the Markov chain  $\{G_m^{(n)}, m \geq 0\}$  and by discussing on the rate of

convergence. Let  $\mathbb{X} = \mathbb{R}$  and let  $\{P_m^{(n)}, m \geq 0\}$  be the Markov chain defined by (12). In particular, for the rest of the section, we consider the mean functional Markov chain  $\{M_m^{(n)}, m \geq 0\}$  defined recursively by

$$M_m^{(n)} = \theta_m \sum_{i=1}^n q_{m,i}^{(n)} Y_{m,i} + (1 - \theta_m) M_{m-1}^{(n)}, \quad m \geq 1, \quad (14)$$

where  $M_0^{(n)} \in \mathbb{R}$  is arbitrary and  $n$  is a given positive integer. From Theorem 4, under the condition  $\int_{\mathbb{R}} \log(1 + |x|) \alpha(dx) < +\infty$ , the Markov chain  $\{M_m^{(n)}, m \geq 0\}$  has the distribution  $\mathcal{M}$  of the random Dirichlet mean  $M$  as the unique invariant measure. It is not restrictive to consider only the chain  $\{M_m^{(n)}, m \geq 0\}$ , since a more general linear functionals  $G$  of a Dirichlet process on an arbitrary Polish space has the same distribution as the mean functional of a Dirichlet process with parameter  $\alpha_g$ , where  $\alpha_g(B) := \alpha(g^{-1}(B))$  for any  $B \in \mathcal{B}$ .

**Theorem 5.** *The Markov chain  $\{M_m^{(n)}, m \geq 0\}$  satisfies the following properties:*

- (i)  $\{M_m^{(n)}, m \geq 0\}$  is a stochastically monotone Markov chain;
- (ii) if further

$$\mathbb{E}[|Y_{1,1}|] = \int_{\mathbb{R}} |x| \alpha_0(dx) < +\infty, \quad (15)$$

then  $\{M_m^{(n)}, m \geq 0\}$  is a geometrically ergodic Markov chain;

- (iii) if the support of  $\alpha$  is bounded then  $\{M_m^{(n)}, m \geq 0\}$  is an uniformly ergodic Markov chain.

Recall that the stochastic monotonicity property of  $\{M_m^{(n)}, m \geq 0\}$  allows to consider exact sampling (see [27]) for  $M$  via  $\{M_m^{(n)}, m \geq 0\}$ .

**Remark 1.** Condition (15) can be relaxed. If the following condition holds

$$\mathbb{E}[|Y_{1,1}|^s] = \int_{\mathbb{R}} |y|^s \alpha_0(dy) < +\infty \quad \text{for some } 0 < s < 1, \quad (16)$$

then the Markov chain  $\{M_m^{(n)}, m \geq 0\}$  is geometrically ergodic. See Appendix A.3 for the proof. If, for instance,  $\alpha_0$  is a Cauchy standard distribution and  $a > 0$ , condition (16) is fulfilled so that  $\{M_m^{(n)}, m \geq 0\}$  will turn out to be geometrically ergodic for any fixed integer  $n$ .

From Theorem 1,  $T^{(n)} := \sum_{1 \leq i \leq n} q_i^{(n)} Y_i = \int_{\mathbb{R}} x (\sum_{1 \leq i \leq n} q_i^{(n)} \delta_{Y_i})(dx)$  converges in distribution to the random Dirichlet mean  $M$  as  $n \rightarrow +\infty$ ; so it is clear that, for a fixed integer  $n$ , the law of  $T^{(n)}$  approximates the law of  $M$  and that the approximation will

be better for  $n$  large. If we reconsider (14), written as

$$M_m^{(n)} = \theta_m T_m^{(n)} + (1 - \theta_m) M_{m-1}^{(n)}, \quad m \geq 1,$$

since the innovation term  $T^{(n)}$  is an approximation in distribution of the limit (as  $m \rightarrow +\infty$ ) r.v.  $M$ , it is intuitive that the rate of convergence will increase as  $n$  gets larger. This is confirmed by the description of small sets  $C^{(n)}$  in (22) (in the proof of Theorem 5). In fact, under (15) or (16), the Markov chain  $\{M_m^{(n)}, m \geq 0\}$  is geometrically or uniformly ergodic since it satisfies a Foster–Lyapunov condition  $PV(x) := \int_{\mathbb{R}} V(y)p(x, dy) \leq \lambda V(x) + b1_{C^{(n)}}(x)$  for a suitable function  $V$ , a small set  $C^{(n)}$  and constants  $b < +\infty$ ,  $0 < \lambda < 1$ . In particular, the small sets  $C^{(n)}$  generalize the corresponding small set  $C$  obtained in Theorem 1 in [15] which can be recovered by setting  $n = 1$ , that is,  $C = [-K(\lambda), K(\lambda)]$  where

$$K(\lambda) := \frac{1 - \lambda + 1/(1+a)\mathbb{E}[|Y_{1,1}|]}{\lambda - a/(1+a)}.$$

Here the size of the small set  $C^{(n)}$  of  $\{M_m^{(n)}, m \geq 0\}$  can be controlled by an additional parameter  $n$ , suggesting the upper bounds of the rate of convergence of the chain  $\{M_m^{(n)}, m \geq 0\}$  depends on  $n$  too.

However, if we would establish an explicit upper bound on the rate of convergence, we would need results like Theorem 2.2 in [30], or Theorems 5.1 and 5.2 in [29]. All these results need a minorization condition to hold for the  $m_0$ th step transition probability  $p_{m_0}^{(n)}(x, A) := \mathbb{P}(M_m^{(n)} \in A | M_0^{(n)} = x)$  for any  $A \in \mathcal{A}$  and  $x \in \mathbb{R}$ , for some positive integer  $m_0$  and all  $x$  in a small set; in particular, if  $\inf_{x \in C^{(n)}} f(z|x) \geq p_0^{(n)}(z)$ , where  $f(z|x)$  is the density of  $p_1^{(n)}(x, \cdot)$  and  $p_0^{(n)}(z)$  is some density such that  $\varepsilon(n) := \int_{\mathbb{R}} p_0^{(n)}(z) dz > 0$ , then

$$p_1^{(n)}(x, A) \geq \varepsilon(n) \int_A \frac{p_0^{(n)}(z)}{\varepsilon(n)} dz = \varepsilon(n) \nu(A), \quad A \in \mathbb{R}, x \in C^{(n)},$$

where  $\nu$  is a probability measure on  $\mathbb{R}$ . In order to check the validity of our intuition that the rate of convergence will increase as  $n$  gets larger, the function  $\varepsilon(n)$  should be increasing with  $n$  in order to prove that the uniform error (when the support of the  $Y_i$ 's is bounded) in total variation between the law of  $M_m^{(n)}$  given  $M_0^{(n)}$  and its limit distribution decreases as  $n$  increases. If  $f_{T^{(n)}}$  is the density of  $T^{(n)}$ , which exists since, conditioning on  $Y_i$ 's,  $T^{(n)}$  is a random Dirichlet mean, then

$$p_1^{(n)}(x, A) = \int_A f(z|x) dz = \int_{\mathbb{R}} \mathbb{P}(\theta_1 y + (1 - \theta_1)x \in A | T^{(n)} = y, M_0^{(n)} = x) f_{T^{(n)}}(y) dy.$$

Therefore, the density function corresponding to  $p_1^{(n)}(x, A)$  is

$$f(z|x) = \frac{1}{B(a, n)} \int_{\mathbb{R}} \frac{(z-x)^{n-1}(y-z)^{a-1}}{(y-x)^{a+n-2}|y-x|} 1_{\{(0,1)\}} \left( \frac{z-x}{y-x} \right) f_{T^{(n)}}(y) dy$$

$$\begin{aligned}
&= \begin{cases} \frac{1}{B(a, n)} \int_{-\infty}^z \frac{(x-z)^{n-1}(z-y)^{a-1}}{(x-y)^{a+n-1}} f_{T^{(n)}}(y) dy, & \text{if } z < x, \\ \frac{1}{B(a, n)} \int_z^{+\infty} \frac{(z-x)^{n-1}(y-z)^{a-1}}{(y-x)^{a+n-1}} f_{T^{(n)}}(y) dy, & \text{if } z > x \end{cases} \\
&= \frac{1}{B(a, n)} \int_0^1 t^{n-2} (1-t)^{a-1} f_{T^{(n)}}\left(\frac{z-(1-t)x}{t}\right) dt.
\end{aligned}$$

Unfortunately, the explicit expression of  $f_{T^{(n)}}$ , which for  $n=1$  reduces to the density of  $\alpha_0$  if it exists, is not simple; from Proposition 5 in [28] for instance, for  $y \in \mathbb{R}$ ,

$$f_{T^{(n)}}(y) = \int_{\mathbb{R}^n} f_{T^{(n)}}(y; y_1, \dots, y_n) F_{(Y_1, \dots, Y_n)}(dy_1, \dots, dy_n),$$

where, when  $y \neq y_i$  for  $i=1, \dots, n$ ,

$$f_{T^{(n)}}(y; y_1, \dots, y_n) = \frac{n-1}{\pi} \int_0^{+\infty} \prod_{j=1}^n \frac{1}{(1+t^2(y_j-y)^2)^{1/2}} \cos\left(\sum_{j=1}^n \arctan(t(y_j-y))\right) dt;$$

here  $F_{(Y_1, \dots, Y_n)}$  is the distribution of  $(Y_1, \dots, Y_n)$  which, by definition, can be recovered by the product rule  $F_{(Y_1, \dots, Y_n)}(y_1, \dots, y_n) = F_{Y_1}(y_1) F_{Y_2|Y_1}(y_2; y_1) \cdots F_{Y_n|Y_1, \dots, Y_{n-1}}(y_n; y_1, \dots, y_{n-1})$  with  $F_1 = A_0$  and

$$F_{Y_j|Y_1, \dots, Y_{j-1}}(y; y_1, \dots, y_{j-1}) = \frac{a}{a+j-1} A_0(y) + \frac{1}{a+j-1} \sum_{i=1}^{j-1} 1_{(-\infty, y]}(y_i).$$

However, some remarks on the asymptotic behavior of  $\varepsilon(n)$  can be made under suitable conditions. Since,

$$\frac{1}{t} f_{T^{(n)}}\left(\frac{z-(1-t)x}{t}\right) \geq f_{T^{(n)}}\left(\frac{z-(1-t)x}{t}\right),$$

if the support of  $Y_i$ 's is bounded (for instance equal to  $[0, 1]$ ) and the derivative of  $f_{T^{(n)}}$  is bounded by some constant  $K$ , then, by Taylor expansion of  $f_{T^{(n)}}$ , we have

$$\begin{aligned}
\varepsilon(n) &= \frac{1}{B(a, n)} \int_0^1 (1-t)^{a-1} t^n \left( \int_0^1 \frac{1}{t} f_{T^{(n)}}(z/t) dz \right) dt \\
&\quad - \frac{1}{B(a, n)} \sup_{x \in C^{(n)}} x \int_0^1 \left( \int_0^1 (1-t)^a t^{n-2} f'_{T^{(n)}}\left(\frac{z_{x,t}}{t}\right) dt \right) dz \\
&\geq \frac{n}{a+n} - K K^{(n)}(\lambda) \int_0^1 (1-t)^a t^{n-2} dt \\
&= \frac{n}{a+n} - \frac{a K K^{(n)}(\lambda)}{n-1}.
\end{aligned} \tag{17}$$

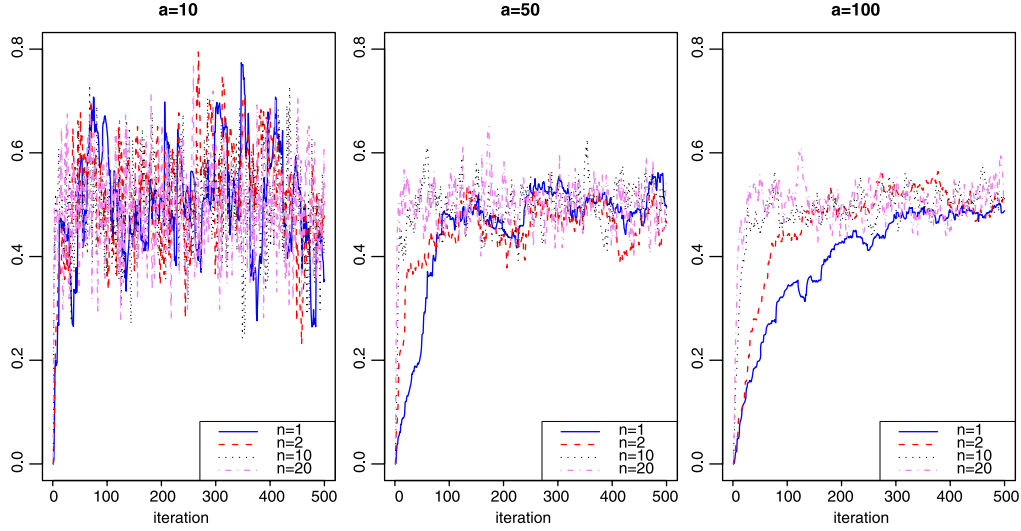
For a large enough  $n_0$ , if we fix  $\lambda$  equal to some positive constant  $C$  which is greater than  $a/(a+n)$  for all  $n > n_0$ , then  $K^{(n)}(\lambda)$  is bounded above by

$$\frac{1 - C + \mathbb{E}[Y_{1,1}]}{C - a/(a+n_0)}.$$

The second term in (17) is negligible with respect to the first term, which increase as  $n$  increases. As we mentioned, when the support of the  $Y_i$ 's is bounded, from Theorem 16.2.4 in [22] it follows that the error in total variation between the  $m$ th transition probability of the Markov chain  $\{M_m^{(n)}, m \geq 0\}$  and the limit distribution  $\mathcal{M}$  is less than  $(1 - \varepsilon(n))^m$ . This error decreases for  $n$  increasing greater than  $n_0$ .

So far we have provided only some qualitative features on the rate of convergence; however, the derivation of the explicit bound of the rate of convergence of  $M_m^{(n)}$  to  $M$  for each fixed  $n$ , via  $p_0^{(n)}$  and  $\varepsilon(n)$ , is still an open problem. Some examples confirm our conjecture that the convergence of the Markov chain  $\{M_m^{(n)}, m \geq 0\}$  improves as  $n$  increases. Nonetheless, we must point out that simulating the innovation term  $T^{(n)}$  for  $n$  larger than 1 will be more computationally expensive, and also that this cost will be increasing as  $n$  increases. In fact, if  $n$  is greater than one,  $2n - 1$  more r.v.s must be drawn at each iteration of the Markov chain ( $n - 1$  more from the Pólya sequence and  $n$  more from the finite-dimensional Dirichlet distribution). Moreover, we compared the total user times of the R function simulating  $\{M_m^{(n)}, m = 0, \dots, 500\}$ . We found that all these times were small, of course depending on  $\alpha_0$ , but not on the total mass parameter  $a$  (all the other values being fixed). The total user times when  $n = 2$  were about 50% greater than those for  $n = 1$ , while they were about 5, 10 and 50 times greater when  $n = 10, 20$  and 100, respectively, for a number of total iterations equal to 500. From the following examples, we found that values of  $n$  between 2 and 20 are a good choice between a fast rate of convergence and a moderate computational cost.

**Example 1.** Let  $\alpha_0$  be a Uniform distribution on  $(0, 1)$  and let  $a$  be the total mass. In this case  $\mathbb{E}[Y_{1,j}] = 1/2$  so that for any fixed integer  $n$ , the chain will be geometrically ergodic; moreover, it can be proved that  $(0, 1)$  is small so that the chain is uniformly ergodic. When  $a = 1$ , [15] showed that the convergence of  $\{M_m, m \geq 0\}$  is very good and there is no need to consider the chain with  $n > 1$ . We consider the cases  $a = 10, 50$  and 100, and for each of them we run the chain for  $n = 1, 2, 10$  and 20. We found that the trace plots do not depend on the initial values. In Figure 1, we give the trace plots of  $M_m^{(n)}$  when  $M_0^{(n)} = 0$ . Observe that convergence improves as  $n$  increases for any fixed value of  $a$ ; however the improvement is more glaring from the graph for large  $a$ . When  $a = 100$  the convergence of the chain for  $n = 1$  seems to occur at about  $m = 350$ , while for  $n = 20$  the convergence is at about a value between 50 and 75. For these values of  $n$ , the total user times to reach convergence was 0.038 seconds for the former, and 0.066 seconds for the latter. Moreover, the total user times to simulate 500 iterations of  $\{M_m^{(n)}, m \geq 0\}$  were 0.05, 0.071, 0.226, 0.429, 2.299 seconds when  $n = 1, 2, 10, 20, 100$ , respectively.



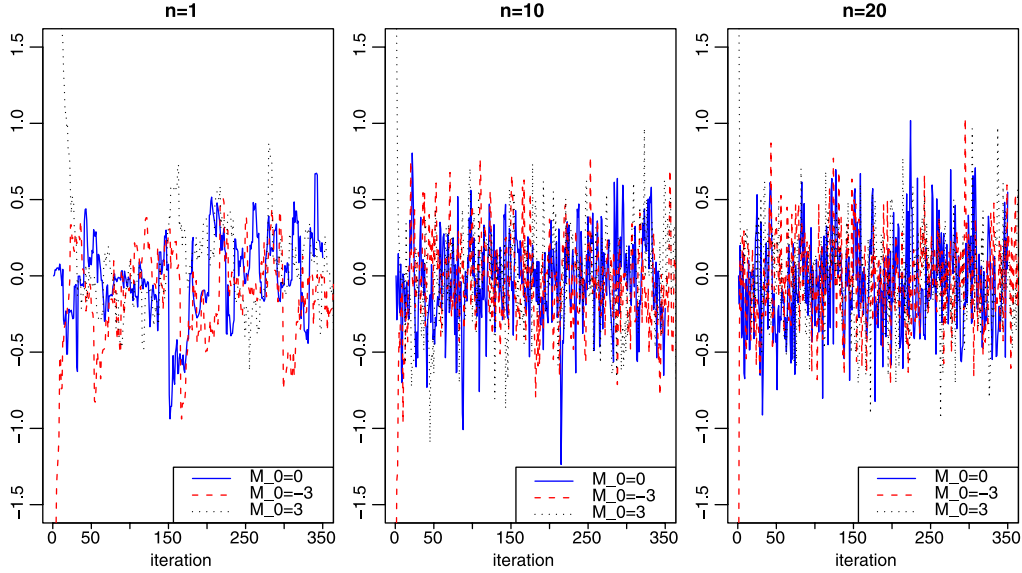
**Figure 1.** Traceplots of the Markov chain  $\{M_m^{(n)}, m \geq 0\}$  with  $\alpha_0$  the Uniform distribution on  $(0, 1)$ ,  $a = 10, 50, 100$  and  $n = 1$  (solid blue line),  $n = 2$  (dashed red line),  $n = 10$  (dotted black line) and  $n = 20$  (dot-dashed violet line).

This behaviour is confirmed in the next example, where the support of the measure  $\alpha$  is assumed to be unbounded.

**Example 2.** Let  $\alpha_0$  be a Gaussian distribution with parameter  $(0, 1)$  and let  $a = 10$ . The behavior of  $M_m, m \geq 0$  has been considered in [15]. Figure 2 displays the trace plots of  $M_m^{(n)}, m \geq 0$  for three different initial values  $(M_0^{(n)} = -3, 0, 3)$ , with  $n = 1, 10, 20$ . Also in this case, it is clear that the convergence improves as  $n$  increases. As far as the total user times are concerned, we drew similar conclusions than in Example 1.

The next is an example in the mixture models context.

**Example 3.** Let us consider a Gaussian kernel  $k(x, \theta)$  with unknown mean  $\theta$  and known variance equal to 1. If we consider the random density  $f(x) = \int_{\mathbb{R}} k(x, \theta) dP(\theta)$ , where  $P$  is a Dirichlet process with parameter  $\alpha$ , then, for any fixed  $x$ ,  $f(x)$  is a random Dirichlet mean. Therefore, if we consider the measure-valued Markov chain  $\{P_m^{(n)}, m \geq 0\}$  defined recursively as in (12), we define a sequence of random densities  $\{f_m^{(n)}(x), m \geq 0\}$ , where  $f_m^{(n)}(x) := \int_{\mathbb{R}} k(x, \theta) dP_m^{(n)}(\theta) = \theta_m \sum_{i=1}^n q_{m,i}^{(n)} k(x, Y_{m,i}) + (1 - \theta_m) f_{m-1}^{(n)}(x)$ . In each panel of Figure 3, we drew  $f_m^{(n)}(x)$  for different values of  $m$  when  $n$  is fixed. In particular, we fixed  $\alpha_0$  to be a Gaussian distribution with parameter  $(0, 1)$ , and let  $a = 100$ ; in this case, since the “variance” of  $P$  is small, the mean density  $E[f](x) = \int_{\mathbb{R}} k(x, \theta) \alpha_0(d\theta)$  (Gaussian with parameter  $(0, 2)$ ) will be very close to the random function  $f(x)$ , so that it can be



**Figure 2.** Traceplots of the Markov chain  $\{M_m^{(n)}, m \geq 0\}$  with  $\alpha_0$  the Gaussian distribution with parameter  $(0, 1)$ ,  $a = 10$ ,  $n = 1, 10, 20$  and  $M_0^{(1)} = -3$  (dashed red line),  $M_0^{(1)} = 0$  (solid blue line) and  $M_0^{(1)} = 3$  (dotted black line).

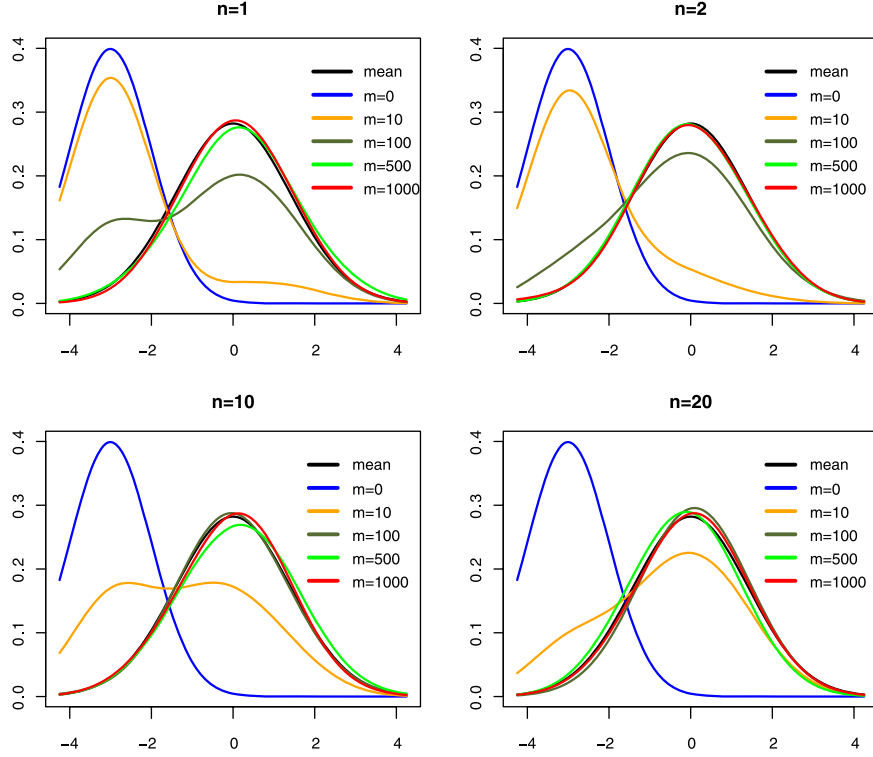
considered an approximation of the “true” density  $f(x)$ . From the plots, it is clear that the convergence improves as  $n$  increases: when  $n = 1$ , only  $f_{1000}^{(1)}(x)$  is close enough to the mean density  $E[f](x)$ , while, if  $n = 20$ ,  $f_{100}^{(20)}(x)$ , as well as the successive iterations, is a good approximation of  $f(x)$ . In any case, observe that even if the “true” density  $f(x)$  is unknown, as when  $a = 1$ , the improvement (as  $n$  increases) is clear as well; see Figure 4, where 5 draws of  $f_m^{(n)}(x)$ ,  $m = 1, 100, 1000$ , are plotted for different values of  $n$ .

## 4. Concluding remarks and developments

The paper [12] constitutes, as far as we know, the first work highlighting the interplay between Bayesian nonparametrics on the one side and the theory of measure-valued Markov chains on the other. In the present paper, we have further studied such interplay by introducing and investigating a new measure-valued Markov chain  $\{Q_n, n \geq 1\}$  defined via exchangeable sequences of r.v.s. Two applications related to Bayesian nonparametrics have been considered: the first gives evidence that  $\{Q_n, n \geq 1\}$  is strictly related to the Newton’s algorithm of the mixture of Dirichlet process model, while the second shows how  $\{Q_n, n \geq 1\}$  can be applied in order to define a generalization of the Feigin–Tweedie Markov chain.

An interesting development consists in investigating whether there are any new applications related to the Feigin–Tweedie Markov chain apart from the well-known application

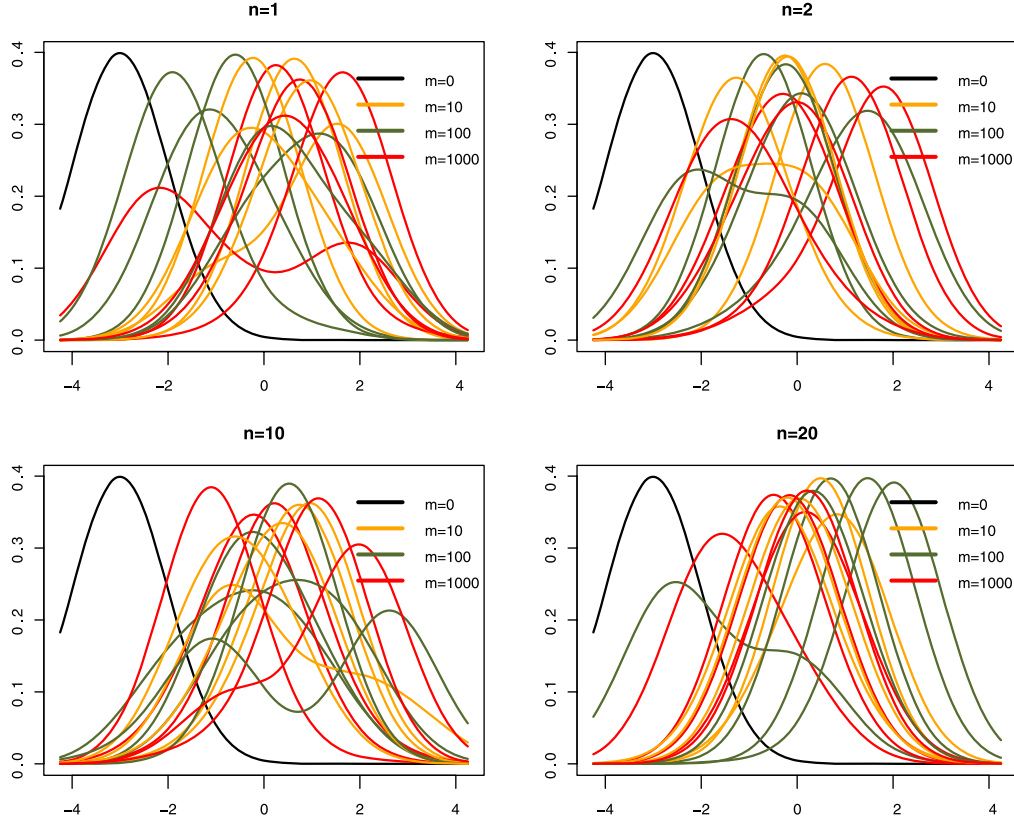




**Figure 3.** Plots of  $f_m^{(n)}$  as in Example 3;  $a = 100$ ,  $\alpha_0 = \mathcal{N}(0, 1)$ ,  $n = 1, 2, 10, 20$  and  $f_0^{(n)} = \mathcal{N}(\cdot; -3, 1)$ .

in the field of functional linear functionals of the Dirichlet process (see, e.g., [14]). The proposed generalization of the Feigin–Tweedie Markov chain represents a large class of measure-valued Markov chains  $\{(P_m^{(n)}, m \geq 0), n \in \mathbb{N}\}$  maintaining all the same properties of the Feigin–Tweedie Markov chain; in other terms, we have increased the flexibility of the Feigin–Tweedie Markov chain via a further parameter  $n \in \mathbb{N}$ . We believe that a number of different interpretations for  $n$  can be investigated in order to extend the applicability of the Feigin–Tweedie Markov chain.

In this respect, an intuitive and simple extension is related to the problem of defining a (bivariate) vector of measure-valued Markov chains  $\{(P_m^{(n_1)}, P_m^{(n_2)}), m \geq 0\}$ , where, for each fixed  $m$ ,  $(P_m^{(n_1)}, P_m^{(n_2)})$  is a vector of dependent random probabilities,  $n_1 < n_2$  being fixed positive integers. Marginally, the two sequences  $\{P_m^{(n_1)}, m \geq 0\}$  and  $\{P_m^{(n_2)}, m \geq 0\}$  are defined via the recursive identity (12); the dependence is achieved using the same Pólya sequence  $(Y_{m,1}^{(n_2)}, \dots, Y_{m,n_1}^{(n_2)}, \dots, Y_{m,n_2}^{(n_2)})$  and assuming dependence in  $(\theta_m^{(n_2)}, \theta_m^{(n_1)})$  or between  $(q_{m,1}^{(n_2)}, \dots, q_{m,n_2}^{(n_2)})$  and  $(q_{m,1}^{(n_1)}, \dots, q_{m,n_1}^{(n_1)})$ . For instance, if, for



**Figure 4.** Plots of 5 draws from  $f_m^{(n)}$  as in Example 3;  $a = 1$ ,  $\alpha_0 = \mathcal{N}(0, 1)$ ,  $n = 1, 2, 10, 20$  and  $f_0^{(n)} = \mathcal{N}(\cdot; -3, 1)$ .

each  $m$ ,  $Z_m, Z_m^{(1)}, \dots, Z_m^{(n_1)}, \dots, Z_m^{(n_2)}$  are independent r.v.s,  $Z_m$  with an Exponential distribution with parameter  $a$ ,  $Z_m^{(i)}$  with an Exponential distribution with parameter 1, we could define  $\theta_m^{(n_2)} := \sum_{i=1}^{n_2} Z_m^{(i)} / (Z_m + \sum_{i=1}^{n_2} Z_m^{(i)})$ ,  $\theta_m^{(n_1)} := \sum_{i=1}^{n_1} Z_m^{(i)} / (Z_m + \sum_{i=1}^{n_1} Z_m^{(i)})$ . Of course, the dependence is related to the difference between  $n_1$  and  $n_2$ . Work on this is ongoing.

## Appendix

### A.1. Weak convergence for the Markov chain $\{Q_n, n \geq 1\}$

A proof of the weak convergence of the sequence  $Q_n = \sum_1^n q_i^{(n)} \delta_{Y_i}$  of r.p.m.s on  $\mathbb{X}$  to a Dirichlet process  $P$  is provided here, when the  $Y_j$ 's are a Pólya sequence with parameter measure  $\alpha$ . The result automatically follows from Theorem 1(ii), but this proof is

interesting *per se*, since we use a combinatorial technique; moreover an explicit expression for the moment of order  $(r_1, \dots, r_k)$  of the  $k$ -dimensional Pólya distribution is obtained.

**Proposition.** *Let  $Q_n$  defined in (5), where  $\{Y_j, j \geq 1\}$  are a Pólya sequence with parameter  $\alpha$ . Then*

$$Q_n \Rightarrow P,$$

where  $P$  is a Dirichlet process on  $\mathbb{X}$  with parameter  $\alpha$ .

**Proof.** By Theorem 4.2 in [18], it is sufficient to prove that for any measurable partition  $B_1, \dots, B_k$  of  $\mathbb{X}$ ,

$$(Q_n(B_1), \dots, Q_n(B_k)) \Rightarrow (P(B_1), \dots, P(B_k)),$$

characterizing the distribution of the limit. For any given measurable partition  $B_1, \dots, B_k$  of  $\mathbb{X}$ , by conditioning on  $Y_1, \dots, Y_n$ , it can be checked that  $(Q_n(B_1), \dots, Q_n(B_{k-1}))$  is distributed according to a Dirichlet distribution with empirical parameter  $(\sum_{1 \leq i \leq n} \delta_{Y_i}(B_1), \dots, \sum_{1 \leq i \leq n} \delta_{Y_i}(B_{k-1}), \sum_{1 \leq i \leq n} \delta_{Y_i}(B_k))$ , and

$$\begin{aligned} \mathbb{P}(\#\{i: Y_i \in B_1\} = j_1, \dots, \#\{i: Y_i \in B_k\} = j_k) \\ = \binom{n}{j_1 \dots j_k} \frac{(\alpha(B_1))_{j_1 \uparrow 1} \dots (\alpha(B_k))_{j_k \uparrow 1}}{(a)_{n \uparrow 1}}, \end{aligned} \quad (18)$$

where  $(j_1, \dots, j_k) \in \mathcal{D}_{k,n}^{(0)}$  with  $\mathcal{D}_{k,n}^{(0)} := \{(j_1, \dots, j_k) \in \{0, \dots, n\}^k : \sum_{1 \leq i \leq k} j_i = n\}$ . For any  $k$ -uple of nonnegative integers  $(r_1, \dots, r_k)$ , we are going to compute the limit, for  $n \rightarrow +\infty$ , of the moment

$$\begin{aligned} \mathbb{E} \left[ (Q_n(B_1))^{r_1} \dots (Q_n(B_{k-1}))^{r_{k-1}} \left( 1 - \sum_{i=1}^{k-1} Q_n(B_i) \right)^{r_k} \right] \\ = \sum_{(j_1, \dots, j_k) \in \mathcal{D}_{k,n}^{(0)}} \binom{n}{j_1 \dots j_k} \frac{(\alpha(B_1))_{j_1 \uparrow 1} \dots (\alpha(B_k))_{j_k \uparrow 1}}{(a)_{n \uparrow 1}} \frac{(j_1)_{r_1 \uparrow 1} \dots (j_k)_{r_k \uparrow 1}}{(n)_{(r_1 + \dots + r_k) \uparrow 1}}, \end{aligned} \quad (19)$$

where in general  $(x)_{n \uparrow \alpha}$  denotes the Pochhammer symbol for the  $n$ th factorial power of  $x$  with increment  $\alpha$ , that is  $(x)_{n \uparrow \alpha} := \prod_{0 \leq i \leq n-1} (x + i\alpha)$ . We will show that, as  $n \rightarrow +\infty$ ,

$$\begin{aligned} \mathbb{E} \left[ (Q_n(B_1))^{r_1} \dots (Q_n(B_{k-1}))^{r_{k-1}} \left( 1 - \sum_{i=1}^{k-1} Q_n(B_i) \right)^{r_k} \right] \\ \rightarrow \mathbb{E} \left[ (P(B_1))^{r_1} \dots (P(B_{k-1}))^{r_{k-1}} \left( 1 - \sum_{i=1}^{k-1} P(B_i) \right)^{r_k} \right], \end{aligned}$$

where  $P$  is a Dirichlet process on  $\mathbb{X}$  with parameter measure  $\alpha$ , that is, the r.v.  $(P(B_1), \dots, P(B_{k-1}))$  has Dirichlet distribution with parameter  $(\alpha(B_1), \dots, \alpha(B_{k-1}))$ ,

$\alpha(B_k)$ ). This will be sufficient to characterize the distribution of the limit  $Q^*$ , because of the boundedness of the support of the limit distribution. First of all, we prove the convergence for  $k = 2$ , which corresponds to the one-dimensional case. In particular, we have

$$\begin{aligned} & \mathbb{E}[(Q_n(B_1))^{r_1}(1 - (Q_n(B_1)))^{r_2}] \\ &= \frac{1}{(n)_{(r_1+r_2)\uparrow 1}} \sum_{(j_1, j_2) \in \mathcal{D}_{2,n}^{(0)}} \binom{n}{j_1, j_2} \frac{(\alpha(B_1))_{j_1\uparrow 1}(\alpha(B_2))_{j_2\uparrow 1}}{(a)_{n\uparrow 1}} (j_1)_{r_1\uparrow 1} (j_2)_{r_2\uparrow 1} \\ &= \frac{1}{(n)_{(r_1+r_2)\uparrow 1}} \sum_{t_1=1}^{r_1} |s(r_1, t_1)| \sum_{s_1=1}^{t_1} S(t_1, s_1) \sum_{t_2=1}^{r_2} |s(r_2, t_2)| \sum_{s_2=1}^{t_2} S(t_2, s_2) \\ & \quad \times \sum_{(j_1, j_2) \in \mathcal{D}_{2,n}^{(0)}} \binom{n}{j_1, j_2} \frac{(\alpha(B_1))_{j_1\uparrow 1}(\alpha(B_2))_{j_2\uparrow 1}}{(a)_{n\uparrow 1}} (j_1)_{s_1\downarrow 1} (j_2)_{s_2\downarrow 1}, \end{aligned}$$

where  $(x)_{n\downarrow 1} := (-1)^{-n}(-x)_{n\uparrow 1}$  and  $s(\cdot, \cdot)$  and  $S(\cdot, \cdot)$  are the Stirling number of the first and second kind, respectively. Let us consider the following numbers, where  $s_1, s_2$  are nonnegative integers and  $n = 1, 2, \dots$ ,

$$C_n^{(s_1, s_2)} := \sum_{(j_1, j_2) \in \mathcal{D}_{2,n}^{(0)}} \binom{n}{j_1, j_2} \frac{(\alpha(B_1))_{j_1\uparrow 1}(\alpha(B_2))_{j_2\uparrow 1}}{(a)_{n\uparrow 1}} (j_1)_{s_1\downarrow 1} (j_2)_{s_2\downarrow 1},$$

and prove they satisfy a recursive relation. In particular,

$$\begin{aligned} C_{n+1}^{(s_1, s_2)} &= \sum_{(j_1, j_2) \in \mathcal{D}_{2,n+1}^{(0)}} \binom{n+1}{j_1, j_2} \frac{(\alpha(B_1))_{j_1\uparrow 1}(\alpha(B_2))_{j_2\uparrow 1}}{(a)_{(n+1)\uparrow 1}} (j_1)_{s_1\downarrow 1} (j_2)_{s_2\downarrow 1} \\ &= \sum_{j_1=0}^n \binom{n+1}{j_1+1} \frac{(\alpha(B_1))_{(j_1+1)\uparrow 1}(\alpha(B_2))_{(n-j_1)\uparrow 1}}{(a)_{(n+1)\uparrow 1}} (j_1+1)_{s_1\downarrow 1} (n-j_1)_{s_2\downarrow 1} \\ &= \frac{(n+1)(\alpha(B_1) + s_1 - 1)}{(a+n)} C_n^{(s_1-1, s_2)} + \frac{n+1}{(a+n)} C_n^{(s_1, s_2)}, \end{aligned}$$

so that the following recursive equation holds

$$C_{n+1}^{(s_1, s_2)} = \frac{(n+1)(\alpha(B_1) + s_1 - 1)}{(a+n)} C_n^{(s_1-1, s_2)} + \frac{n+1}{(a+n)} C_n^{(s_1, s_2)}. \quad (20)$$

Therefore, starting from  $C_n^{(0,0)} = 1$ ,  $C_n^{(0,1)} = n\alpha(B_2)/a$ , we have

$$C_n^{(1,1)} = \sum_{i=0}^{n-1} \frac{(i+1)\alpha(B_1)}{(a+i)} C_i^{(0,1)} \prod_{j=i+1}^{n-1} \frac{j+1}{a+j}$$

$$\begin{aligned}
&= \frac{\Gamma(n+1)\alpha(B_1)}{\Gamma(a+n)} \sum_{i=0}^{n-1} \frac{\Gamma(a+i)}{\Gamma(i+1)} C_i^{(0,1)} \\
&= \frac{\Gamma(n+1)\alpha(B_1)}{\Gamma(a+n)} \sum_{i=0}^{n-1} \frac{\Gamma(a+i)}{\Gamma(i+1)} \frac{i\alpha(B_2)}{a} = \frac{\alpha(B_1)\alpha(B_2)(n)_{2\downarrow 1}}{(a)_{2\uparrow 1}},
\end{aligned}$$

and by (20) we obtain  $C_n^{(s_1, s_2)} = (\alpha(B_1))_{s_1\uparrow 1} (\alpha(B_2))_{s_2\uparrow 1} (n)_{(s_1+s_2)\downarrow 1} / (a)_{(s_1+s_2)\uparrow 1}$ . Thus,

$$\begin{aligned}
&\lim_{n \rightarrow +\infty} \mathbb{E}[(H_n(B_1))^{r_1} (1 - (H_n(B_1)))^{r_2}] \\
&= \sum_{t_1=0}^{r_1} |s(r_1, t_1)| \sum_{s_1=0}^{t_1} S(t_1, s_1) \sum_{t_2=0}^{r_2} |s(r_2, t_2)| \sum_{s_2=0}^{t_2} S(t_2, s_2) \lim_{n \rightarrow +\infty} \frac{C_n^{(s_1, s_2)}}{(n)_{(r_1+r_2)\uparrow 1}} \\
&= |s(r_1, r_1)| S(r_1, r_1) |s(r_2, r_2)| S(r_2, r_2) \frac{(\alpha(B_1))_{r_1\uparrow 1} (\alpha(B_2))_{r_2\uparrow 1}}{(a)_{(r_1+r_2)\uparrow 1}} \\
&= \frac{(\alpha(B_1))_{r_1\uparrow 1} (\alpha(B_2))_{r_2\uparrow 1}}{(a)_{(r_1+r_2)\uparrow 1}}.
\end{aligned}$$

The last expression is exactly  $\mathbb{E}[(P(B_1))^{r_1} (1 - P(B_1))^{r_2}]$ , where  $P(B_1)$  has Beta distribution with parameter  $(\alpha(B_1), \alpha(B_2))$ .

This proof can be easily generalized to the case  $k > 2$ . Analogously to the one-dimensional case, we can write

$$\begin{aligned}
&\mathbb{E} \left[ (Q_n(B_1))^{r_1} \cdots (Q_n(B_{k-1}))^{r_{k-1}} \left( 1 - \sum_{i=1}^{k-1} Q_n(B_i) \right)^{r_k} \right] \\
&= \frac{1}{(n)_{(r_1+\cdots+r_k)\uparrow 1}} \sum_{(j_1, \dots, j_k) \in \mathcal{D}_{k,n}^{(0)}} \binom{n}{j_1 \cdots j_k} \frac{(\alpha(B_1))_{j_1\uparrow 1} \cdots (\alpha(B_k))_{j_k\uparrow 1}}{(a)_{n\uparrow 1}} (j_1)_{r_1\uparrow 1} \cdots (j_k)_{r_k\uparrow 1} \\
&= \frac{1}{(n)_{(r_1+\cdots+r_k)\uparrow 1}} \sum_{t_1=0}^{r_1} |s(r_1, t_1)| \sum_{s_1=0}^{t_1} S(t_1, s_1) \cdots \sum_{t_k=0}^{r_k} |s(r_k, t_k)| \sum_{s_k=0}^{t_k} S(t_k, s_k) \\
&\quad \times \sum_{(j_1, \dots, j_k) \in \mathcal{D}_{k,n}^{(0)}} \binom{n}{j_1 \cdots j_k} \frac{(\alpha(B_1))_{j_1\uparrow 1} \cdots (\alpha(B_k))_{j_k\uparrow 1}}{(a)_{n\uparrow 1}} (j_1)_{s_1\downarrow 1} \cdots (j_k)_{s_k\downarrow 1}
\end{aligned}$$

and, as before, define

$$C_n^{(s_1, \dots, s_k)} := \sum_{(j_1, \dots, j_k) \in \mathcal{D}_{k,n}^{(0)}} \binom{n}{j_1 \cdots j_k} \frac{(\alpha(B_1))_{j_1\uparrow 1} \cdots (\alpha(B_k))_{j_k\uparrow 1}}{(a)_{n\uparrow 1}} \times (j_1)_{s_1\downarrow 1} \cdots (j_k)_{s_k\downarrow 1}$$

and prove they satisfy a recursive relation. Observe that  $C_n^{(s_1, \dots, s_k)} / (n)_{(r_1+\cdots+r_k)\uparrow 1}$  is the moment of order  $(r_1, \dots, r_k)$  of the  $k$ -dimensional Pólya distribution by definition.

Therefore, for  $k > 2$

$$\begin{aligned}
C_n^{(s_1, \dots, s_k)} &= \sum_{j_k=0}^n \binom{n}{j_k} \frac{(\alpha(B_k))_{j_k \uparrow 1} (a - \alpha(B_k))_{(n-j_k) \uparrow 1} (j_k)_{s_k \downarrow 1}}{(a)_{n \uparrow 1}} \\
&\quad \times \sum_{(j_1, \dots, j_{k-1}) \in \mathcal{D}_{k-1, n-j_k}^{(0)}} \binom{n-j_k}{j_1 \cdots j_{k-1}} \frac{(\alpha(B_1))_{j_1 \uparrow 1} \cdots (\alpha(B_{k-1}))_{j_{k-1} \uparrow 1}}{(a - \alpha(B_k))_{(n-j_k) \uparrow 1}} \\
&\quad \times (j_1)_{s_1 \downarrow 1} \cdots (j_{k-1})_{s_{k-1} \downarrow 1} \\
&= \sum_{j_k=0}^n \binom{n}{j_k} \frac{(\alpha(B_k))_{j_k \uparrow 1} (a - \alpha(B_k))_{(n-j_k) \uparrow 1} (j_k)_{s_k \downarrow 1}}{(a)_{n \uparrow 1}} \\
&\quad \times \frac{(\alpha(B_1))_{s_1 \uparrow 1} \cdots (\alpha(B_{k-1}))_{s_{k-1} \uparrow 1} (n-j_k)_{(s_1 + \cdots + s_{k-1}) \downarrow 1}}{(a - \alpha(B_k))_{(s_1 + \cdots + s_{k-1}) \uparrow 1}},
\end{aligned}$$

where the last equality follows by induction hypothesis (we already proved the base case  $k = 2$  in (20)). Then, following the same steps of the one-dimensional case, we can recover a recursive equation for  $C_n^{(s_1, \dots, s_k)}$ ,

$$C_{n+1}^{(s_1, \dots, s_k)} = \frac{(n+1)(\alpha(B_k) + s_k - 1)}{a+n} C_n^{(s_1, \dots, s_{k-1}, s_k-1)} + \frac{n+1}{a+n} C_n^{(s_1, \dots, s_k)}. \quad (21)$$

Starting from  $C_n^{(0, \dots, 0)} = 1$ ,  $C_n^{(0, \dots, 0, 1, 0, \dots, 0)} = n\alpha(B_j)/a$  and

$$\begin{aligned}
C_n^{(1, \dots, 1)} &= \sum_{i=0}^{n-1} \frac{(i+1)\alpha(B_k)}{(a+i)} C_i^{(1, \dots, 1, 0)} \prod_{j=i+1}^{n-1} \frac{j+1}{a+j} \\
&= \frac{\Gamma(n+1)}{\Gamma(a+n)} \frac{(\alpha(B_1))_{s_1 \uparrow 1} \cdots (\alpha(B_{k-1}))_{s_{k-1} \uparrow 1} \alpha(B_k)}{(a - \alpha(B_k))_{(s_1 + \cdots + s_{k-1}) \uparrow 1}} \sum_{i=0}^{n-1} \frac{\Gamma(a+i)}{\Gamma(1+i)} C_i^{(1, \dots, 1, 0)} \\
&= \frac{\Gamma(n+1)}{\Gamma(a+n)} \frac{(\alpha(B_1))_{s_1 \uparrow 1} \cdots (\alpha(B_{k-1}))_{s_{k-1} \uparrow 1} \alpha(B_k)}{(a - \alpha(B_k))_{(s_1 + \cdots + s_{k-1}) \uparrow 1}} \\
&\quad \times (-1)^{k+1} \sum_{i=0}^{n-1} \frac{\Gamma(a+i)}{\Gamma(1+i)} \frac{(a - \alpha(B_k))_{i \uparrow 1} (-i)_{(k-1) \uparrow 1}}{(a)_{i \uparrow 1}} \\
&\quad \times \frac{\Gamma(-a + \alpha(B_k) - i + 1) \Gamma(-k - a + 2)}{\Gamma(-a - i + 1) \Gamma(-a + \alpha(B_k) + 2 - k)} \\
&= \frac{\alpha(B_1) \cdots \alpha(B_k) (n)_{k \downarrow 1}}{(a)_{(k \uparrow 1)}},
\end{aligned}$$

by repeated application of (21), we obtain

$$C_n^{(s_1, \dots, s_k)} = \frac{(\alpha(B_1))_{s_1 \uparrow 1} \cdots (\alpha(B_k))_{s_k \uparrow 1} (n)_{(s_1 + \cdots + s_k) \downarrow 1}}{(a)_{(s_1 + \cdots + s_k) \uparrow 1}}.$$

Thus,

$$\begin{aligned}
& \lim_{n \rightarrow +\infty} \mathbb{E} \left[ (Q_n(B_1))^{r_1} \cdots (Q_n(B_{k-1}))^{r_{k-1}} \left( 1 - \sum_{i=1}^{k-1} Q_n(B_i) \right)^{r_p} \right] \\
&= \sum_{t_1=0}^{r_1} |s(r_1, t_1)| \sum_{s_1=0}^{t_1} S(t_1, s_1) \cdots \sum_{t_k=0}^{r_k} |s(r_k, t_k)| \sum_{s_k=0}^{t_k} S(t_k, s_k) \lim_{n \rightarrow +\infty} \frac{C_n^{(s_1, \dots, s_k)}}{(n)_{(r_1 + \dots + r_k) \uparrow 1}} \\
&= |s(r_1, r_1)| S(r_1, r_1) \cdots |s(r_k, r_k)| S(r_k, r_k) \frac{(\alpha(B_1))_{r_1 \uparrow 1} \cdots (\alpha(B_k))_{r_k \uparrow 1}}{(a)_{(r_1 + \dots + r_k) \uparrow 1}} \\
&= \frac{(\alpha(B_1))_{r_1 \uparrow 1} \cdots (\alpha(B_k))_{r_k \uparrow 1}}{(a)_{(r_1 + \dots + r_k) \uparrow 1}} \\
&= \mathbb{E} \left[ (P(B_1))^{r_1} \cdots (P(B_{k-1}))^{r_{k-1}} \left( 1 - \sum_{i=1}^{k-1} P(B_i) \right)^{r_p} \right],
\end{aligned}$$

where  $P$  is a Dirichlet process with parameter  $\alpha$ . □

## A.2. Solution of the distributional equation

Here, we provide an alternative proof for the solution of the distributional equation (3) introduced by [11].

**Theorem.** *For any fixed integer  $n \geq 1$ , the distributional equation*

$$P^{(n)} \stackrel{d}{=} \theta \sum_{i=1}^n q_i^{(n)} \delta_{Y_i} + (1 - \theta) P^{(n)}$$

*has the Dirichlet process with parameter  $\alpha$  as its unique solution, assuming the independence between  $P^{(n)}$ ,  $\theta, (q_1^{(n)}, \dots, q_n^{(n)})$  and  $(Y_1, \dots, Y_n)$  in the right-hand side.*

**Proof.** From Skorohod's theorem, it follows that there exist  $n$  independent r.v.s  $\xi_1, \dots, \xi_n$  such that  $\xi_i$  has Beta distribution with parameter  $(1, n - i)$  for  $i = 1, \dots, n$  and  $q_1^{(n)} = \xi_1$  and  $q_i^{(n)} = \xi_i \prod_{1 \leq j \leq i-1} (1 - \xi_j)$  for  $i = 2, \dots, n$ ; in particular, by a simple transformation of r.v.s, it follows that  $(q_1^{(n)}, \dots, q_n^{(n)})$  is distributed according to a Dirichlet distribution function with parameter  $(1, \dots, 1)$ . Further, since  $\xi_n = 1$  a.s., then  $\sum_{1 \leq i \leq n} q_i^{(n)} = 1$  a.s. and it can be verified by induction that

$$1 - \sum_{i=1}^j q_i^{(n)} = \prod_{i=1}^j (1 - \xi_i), \quad j = 1, \dots, n-1.$$

Let  $B_1, \dots, B_k$  be a measurable partition of  $\mathbb{X}$ . We first prove that conditionally on  $Y_1, \dots, Y_n$ , the finite dimensional distribution of the r.p.m.  $\sum_{1 \leq i \leq n} q_i^{(n)} \delta_{Y_i}$  is the Dirichlet distribution with the empirical parameter  $(\sum_{1 \leq i \leq n} \delta_{Y_i}(B_1), \dots, \sum_{1 \leq i \leq n} \delta_{Y_i}(B_k))$ . Actually, since

$$\begin{aligned} & \left( \left( \sum_{i=1}^n q_i^{(n)} \delta_{Y_i} \right) (\cdot, B_1), \dots, \left( \sum_{i=1}^n q_i^{(n)} \delta_{Y_i} \right) (\cdot, B_k) \right) \\ &= \left( \sum_{i=1}^n q_i^{(n)} \delta_{Y_i}(B_1), \dots, \sum_{i=1}^n q_i^{(n)} \delta_{Y_i}(B_k) \right) \\ &= \left( \sum_{i: Y_i \in B_1} q_i^{(n)}, \dots, \sum_{i: Y_i \in B_k} q_i^{(n)} \right), \end{aligned}$$

then, conditionally on the r.v.s  $Y_1, \dots, Y_n$ , the r.v.  $(\sum_{i: Y_i \in B_1} q_i^{(n)}, \dots, \sum_{i: Y_i \in B_k} q_i^{(n)})$  is distributed according to a Dirichlet distribution with parameter  $(n_1, \dots, n_k)$ , where  $n_i = \sum_{1 \leq j \leq n} \delta_{Y_j}(B_i)$  for  $i = 1, \dots, k$ . Conditionally on  $Y_1, \dots, Y_n$ , the finite dimensional distributions of the right-hand side of (3) are Dirichlet with updated parameter  $((\alpha(B_1) + \sum_{1 \leq i \leq n} \delta_{Y_i}(B_1), \dots, \alpha(B_k) + \sum_{1 \leq i \leq n} \delta_{Y_i}(B_k)))$ . This argument verifies that the Dirichlet process with parameter  $\alpha$  satisfies the distributional equation (3). This solution is unique by Lemma 3.3 in [33] (see also [37], Section 1).  $\square$

### A.3. Proofs of the theorems in Section 3

**Proof of Theorem 3.** The proof is based on the “standard” result that properties (e.g., weak convergence) of sequences of r.p.m.s can be proved via analogous properties of the sequences of their linear functionals.

First of all, we prove that if  $g: \mathbb{X} \mapsto \mathbb{R}$  is a bounded and continuous function, then  $\{G_m^{(n)}, m \geq 0\}$  with  $G_m^{(n)} := \int_{\mathbb{X}} g(x) P_m^{(n)}(dx)$  is a Markov chain on  $\mathbb{R}$  with unique invariant measure  $\Pi_g$ . From (12), it follows that  $\{G_m^{(n)}, m \geq 1\}$  is a Markov chain on  $\mathbb{R}$  restricted to the compact set  $[-\sup_{\mathbb{X}} |g(x)|, \sup_{\mathbb{X}} |g(x)|]$  and it has at least one finite invariant measure if it is a weak Feller Markov chain. In fact, for a fixed  $y \in \mathbb{R}$

$$\begin{aligned} & \liminf_{x \rightarrow x^*} \mathbb{P}(G_m^{(n)} \leq y | G_{m-1}^{(n)} = x) \\ &= \liminf_{x \rightarrow x^*} \mathbb{P} \left( \theta_m \sum_{i=1}^n q_{m,i}^{(n)} g(Y_{m,i}) \leq y - x(1 - \theta_m) \right) \\ &\geq \int_{(0,1)} \liminf_{x \rightarrow x^*} \mathbb{P} \left( \sum_{i=1}^n q_{m,i}^{(n)} g(Y_{m,i}) \leq \frac{y - x(1 - z)}{z} \right) \mathbb{P}(\theta_m \in dz) \end{aligned}$$



$$\begin{aligned}
&= \int_{(0,1)} \mathbb{P} \left( \sum_{i=1}^n q_{m,i}^{(n)} g(Y_{m,i}) \leq \frac{y - x^*(1-z)}{z} \right) \mathbb{P}(\theta_m \in dz) \\
&= \mathbb{P}(G_m^{(n)} \leq y | G_{m-1}^{(n)} = x^*),
\end{aligned}$$

since the distribution of  $\sum_{1 \leq i \leq n} q_{m,i}^{(n)} g(Y_{m,i})$  has at most a countable numbers of atoms and  $\theta_m$  is absolutely continuous. From Proposition 4.3 in [36], if we show that  $\{G_m^{(n)}, m \geq 0\}$  is  $\phi$ -irreducible for a finite measure  $\phi$ , then the Markov chain is positive recurrent and the invariant measure  $\Pi_g$  is unique. Let us consider the following event  $E := \{Y_{1,1} = Y_{1,2} = \dots = Y_{1,n}\}$ . Then for a finite measure  $\phi$  we have to prove that if  $\phi(A) > 0$ , then  $\mathbb{P}(G_1^{(n)} \in A | G_0^{(n)}) > 0$  for any  $G_0^{(n)}$ . We observe that

$$\begin{aligned}
\mathbb{P}(G_1^{(n)} \in A | G_0^{(n)}) &= \mathbb{P}(G_1^{(n)} \in A | G_0^{(n)}, E) \mathbb{P}(E | G_0^{(n)}) + \mathbb{P}(G_1^{(n)} \in A | G_0^{(n)}, E^c) \mathbb{P}(E^c | G_0^{(n)}) \\
&\geq \mathbb{P}(G_1^{(n)} \in A | G_0^{(n)}, E) \mathbb{P}(E).
\end{aligned}$$

Therefore, since  $\mathbb{P}(E) > 0$ , using the same argument in Lemma 2 in [12], we conclude that  $\mathbb{P}(G_1^{(n)} \in A | G_0^{(n)}, E) > 0$  for a suitable measure  $\phi$  such that  $\phi(A) > 0$ . Finally, we prove the aperiodicity of  $\{G_m^{(n)}, m \geq 0\}$  by contradiction. If the chain is periodic with period  $d > 1$ , then exist  $d$  disjoint sets  $D_1, \dots, D_d$  such that  $\mathbb{P}(G_m^{(n)} \in D_{i+1} | G_{m-1}^{(n)} = x) = 1$  for all  $x \in D_i$  and for  $i = 1, \dots, d-1$ . This implies  $\mathbb{P}(z \sum_{i=1}^n q_{m,i}^{(n)} g(Y_{m,i}) + (1-z)x \in D_{i+1}) = 1$  for almost every  $z$  with respect to the Lebesgue measure restricted to  $(0, 1)$ . Thus,  $\mathbb{P}(\sum_{i=1}^n q_{m,i}^{(n)} g(Y_{m,i}) \in D_{i+1}) = 1$  for  $i = 0, \dots, d-1$ . For generic  $\alpha$  and  $g$ , this is in contradiction with the assumption  $d > 1$ . By Theorem 13.3.4(ii) in [22],  $G_m^{(n)}$  converges in distribution for  $\Pi_g$ -almost all starting points  $G_0^{(n)}$ . In particular,  $\{G_m^{(n)}, m \geq 0\}$  converges weakly for  $\Pi_g$ -almost all starting points  $G_0^{(n)}$ .

From the arguments above, it follows that, for all  $g$  bounded and continuous, there exists a r.v.  $G$  such that  $G_m^{(n)} \Rightarrow G$  as  $m \rightarrow +\infty$  for  $\Pi_g$ -almost all starting points  $G_0^{(n)}$ . Therefore, for Lemma 5.1 in [18] there exists a r.p.m.  $P$  such that  $P_m^{(n)} \Rightarrow P$  as  $m \rightarrow +\infty$  and  $G \stackrel{d}{=} \int_{\mathbb{X}} g(x) P(\cdot, dx)$  for all  $g \in C(\mathbb{R})$ . This implies that the law of  $P$  is an invariant measure for the Markov chain  $\{P_m^{(n)}, m \geq 0\}$ . Then, as  $m \rightarrow +\infty$ ,

$$\int_{\mathbb{X}} g(x) dP_m^{(n)}(\cdot, dx) \Rightarrow \int_{\mathbb{X}} g(x) P(\cdot, dx)$$

and the limit is unique for any  $g \in C(\mathbb{R})$ . Since for any random measure  $\zeta_1$  and  $\zeta_2$  we know that  $\zeta_1 \stackrel{d}{=} \zeta_2$  if and only if  $\int_{\mathbb{X}} g(x) \zeta_1(\cdot, dx) = \int_{\mathbb{X}} g(x) \zeta_2(\cdot, dx)$  for any  $g \in C(\mathbb{R})$  (see Theorem 3.1. in [18]), the invariant measure for the Markov chain  $\{P_m^{(n)}, m \geq 0\}$  is unique. By the definition of  $\{P_m^{(n)}, m \geq 0\}$ , it is straightforward to show that the limit  $P$  must satisfy (3) so that  $P$  is the Dirichlet process with parameter  $\alpha$ .  $\square$

**Proof of Theorem 4.** The proof is a straightforward adaptation of the proof of Theorem 2 in [12], using

$$\begin{aligned} f(G_1^{(n)}) &= \log \left( 1 + \left| \theta_1 \sum_{i=1}^n q_{1,i}^{(n)} g(Y_{1,i}) + (1 - \theta_1) G_0^{(n)} \right| \right) \\ &\leq \sum_{i=1}^n \log(1 + |g(Y_{1,i})|) + \log(1 + (1 - \theta_1) |G_0^{(n)}|) \end{aligned}$$

instead of their inequality (8).  $\square$

**Proof of Theorem 5.** As regards (i), given the definition of stochastically monotone Markov chain, we have that for  $z_1 < z_2$ ,  $s \in \mathbb{R}$ ,

$$\begin{aligned} p_1^{(n)}(z_1, (-\infty, s)) &= \mathbb{P} \left( \theta_1 \sum_{i=1}^n q_{1,i}^{(n)} Y_{1,i} + (1 - \theta_1) z_1 < s \right) \\ &\geq \mathbb{P} \left( \theta_1 \sum_{i=1}^n q_{1,i}^{(n)} Y_{1,i} + (1 - \theta_1) z_2 < a \right) \\ &= p_1^{(n)}(z_2, (-\infty, s)). \end{aligned}$$

As far as (ii) is concerned, we first prove that, under condition (15), the Markov chain  $\{M_m^{(n)}, m \geq 0\}$  satisfies the Foster–Lyapunov condition for the function  $V(x) = 1 + |x|$ . This property implies the geometric ergodicity of the  $\{M_m^{(n)}, m \geq 0\}$  (see [22], Chapter 15). We have

$$\begin{aligned} pV(x) &= \int_{\mathbb{X}} (1 + |y|) p(x, dy) = 1 + \mathbb{E} \left[ \left| \theta_1 \sum_{i=1}^n q_{1,i}^{(n)} Y_{1,i} + (1 - \theta_1) x \right| \right] \\ &\leq 1 + \mathbb{E}[\theta_1] \sum_{i=1}^n \mathbb{E}[|q_{1,i}^{(n)} Y_{1,i}|] + |x| \mathbb{E}[1 - \theta_1] \\ &\leq 1 + \frac{n}{n+a} \sum_{i=1}^n \mathbb{E}[|q_{1,i}^{(n)} Y_{1,i}|] + \frac{a}{n+a} |x| = 1 + \frac{n}{n+a} \mathbb{E}[|Y_{1,1}|] + \frac{a}{n+a} |x|. \end{aligned}$$

Therefore, we are looking for the small set  $C^{(n)}$  such that the Foster–Lyapunov condition holds, that is, a small set  $C^{(n)}$  such that

$$1 + \frac{n}{n+a} \mathbb{E}[|Y_{1,1}|] + \frac{a}{n+a} |x| \leq \lambda(1 + |x|) + b 1_{C^{(n)}}(x) \quad (22)$$

for some constant  $b < +\infty$  and  $0 < \lambda < 1$ . If  $C^{(n)} = [-K^{(n)}(\lambda), K^{(n)}(\lambda)]$ , where

$$K^{(n)}(\lambda) := \frac{1 - \lambda + n/(n+a) \mathbb{E}[|Y_{1,1}|]}{\lambda - a/(n+a)},$$

then, condition (22) holds for all

$$\lambda \in \left( \frac{a}{n+a}, 1 \right), \quad b \geq 1 - \lambda + \frac{n}{n+a} \mathbb{E}[|Y_{1,1}|].$$

As in the proof of Theorem 3, we can prove that the Markov chain  $\{M_m^{(n)}, m \geq 0\}$  is weak Feller; then, since  $C^{(n)}$  is a compact set, it is a small set (see [35]). As regards (iii), the proof follows by standard arguments. See, for instance, the proof of Theorem 1 in [15].  $\square$

**Proof of Remark 1.** As we have already mentioned, the geometric ergodicity follows if a Foster–Lyapunov condition holds. Let  $V(x) = 1 + |x|^s$ ; then, if  $\mathbb{E}[(1 + |\sum_{1 \leq i \leq n} q_{1,i}^{(n)} Y_{1,i}|)^s] < +\infty$ , it is straightforward to prove that the Foster–Lyapunov condition  $PV(x) \leq \lambda V(x) + b1_{\tilde{C}^{(n)}}(x)$  holds for some constant  $b < +\infty$ , and  $\lambda$  such that

$$\mathbb{E}[(1 - \theta_1)^s] = \frac{\Gamma(a+s)\Gamma(a+n)}{\Gamma(a)\Gamma(a+s+n)} < \lambda < 1$$

and for some compact set  $\tilde{C}^{(n)}$ . Of course (16) implies  $\mathbb{E}[(1 + |\sum_{1 \leq i \leq n} q_{1,i}^{(n)} Y_{1,i}|)^s] < +\infty$ ; in fact, conditioning on the random number  $N$  of distinct values  $Y_{1,1}^*, \dots, Y_{1,N}^*$  in  $Y_{1,1}, \dots, Y_{1,n}$ ,  $1 \leq N \leq n$ , we have

$$\left| \sum_{i=1}^n q_{1,i}^{(n)} Y_{1,i} \right| \leq \sum_{i=1}^N \tilde{q}_{1,i}^{(n)} |Y_{1,i}^*| \leq \max\{|Y_{1,1}^*|, \dots, |Y_{1,N}^*|\}.$$

Since  $\{Y_{1,1}^*, \dots, Y_{1,N}^*\}$  are independent and identically distributed according to  $\alpha_0$ , then

$$\mathbb{E} \left[ \left| \sum_{i=1}^n q_{1,i}^{(n)} Y_{1,i} \right|^s \right] \leq \int_0^{+\infty} y^s N(A_0(y))^{N-1} \alpha_0(dy) \leq N \int_0^{+\infty} y^s \alpha_0(dy) \leq n \mathbb{E}[|Y_{1,1}|^s] < +\infty,$$

where  $A_0$  is the distribution corresponding to the probability measure  $\alpha_0$ , and this is equivalent to  $\mathbb{E}[(1 + |\sum_{1 \leq i \leq n} q_{1,i}^{(n)} Y_{1,i}|)^s] < +\infty$ .  $\square$

## Acknowledgements

The authors are very grateful to Patrizia Berti and Pietro Rigo who suggested the proof of Theorem 2, and to Eugenio Regazzini for helpful discussions. The authors are also grateful to an Associate Editor and a referee for comments that helped to improve the presentation. The second author was partially supported by MiUR Grant 2006/134525.

## References

- [1] BERTI, P., PRATELLI, L. and RIGO, P. (2006). Almost sure weak convergence of random probability measures. *Stochastics* **78** 91–97. [MR2236634](#)

- [2] BLACKWELL, D. and MACQUEEN, J.B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1** 353–355. [MR0362614](#)
- [3] CHOW, Y.S. and TEICHER, H. (1997). *Probability Theory: Independence, Interchangeability, Martingales*, 3rd ed. *Springer Texts in Statistics*. New York: Springer. [MR1476912](#)
- [4] CIFARELLI, D.M. and REGAZZINI, E. (1990). Distribution functions of means of a Dirichlet process. *Ann. Statist.* **18** 429–442. [MR1041402](#)
- [5] DAWSON, D.A. (1993). Measure-valued Markov processes. In *École D'Été de Probabilités de Saint-Flour XXI—1991. Lecture Notes in Math.* **1541** 1–260. Berlin: Springer. [MR1242575](#)
- [6] DEL MORAL, P. (2004). *Feynman–Kac Formulae. Genealogical and Interacting Particle Systems with Applications. Probability and Its Applications (New York)*. New York: Springer. [MR2044973](#)
- [7] DOSS, H. and SELKE, T. (1982). The tails of probabilities chosen from a Dirichlet prior. *Ann. Statist.* **10** 1302–1305. [MR0673666](#)
- [8] ERHARDSSON, T. (2008). Non-parametric Bayesian inference for integrals with respect to an unknown finite measure. *Scand. J. Statist.* **35** 369–384. [MR2418747](#)
- [9] ETHERIDGE, A.M. (2000). *An Introduction to Superprocesses. University Lecture Series* **20**. Providence, RI: Amer. Math. Soc. [MR1779100](#)
- [10] ETHIER, S.N. and KURTZ, T.G. (1993). Fleming-Viot processes in population genetics. *SIAM J. Control Optim.* **31** 345–386. [MR1205982](#)
- [11] FAVARO, S. and WALKER, S.G. (2008). A generalized constructive definition for the Dirichlet process. *Statist. Probab. Lett.* **78** 2836–2838. [MR2465128](#)
- [12] FEIGIN, P.D. and TWEEDIE, R.L. (1989). Linear functionals and Markov chains associated with Dirichlet processes. *Math. Proc. Cambridge Philos. Soc.* **105** 579–585. [MR0985694](#)
- [13] GHOSH, J.K. and TOKDAR, S.T. (2006). Convergence and consistency of Newton's algorithm for estimating mixing distribution. In *Frontiers in Statistics* (J. FAN and H. KOUL, eds.) 429–443. London: Imp. Coll. Press. [MR2326012](#)
- [14] GUGLIELMI, A., HOLMES, C.C. and WALKER, S.G. (2002). Perfect simulation involving functionals of a Dirichlet process. *J. Comput. Graph. Statist.* **11** 306–310. [MR1938137](#)
- [15] GUGLIELMI, A. and TWEEDIE, R.L. (2001). Markov chain Monte Carlo estimation of the law of the mean of a Dirichlet process. *Bernoulli* **7** 573–592. [MR1849368](#)
- [16] HANNUM, R.C., HOLLANDER, M. and LANGBERG, N.A. (1981). Distributional results for random functionals of a Dirichlet process. *Ann. Probab.* **9** 665–670. [MR0630318](#)
- [17] JARNER, S.F. and TWEEDIE, R.L. (2002). Convergence rates and moments of Markov chains associated with the mean of Dirichlet processes. *Stochastic Process. Appl.* **101** 257–271. [MR1931269](#)
- [18] KALLENBERG, O. (1983). *Random Measures*, 3rd ed. Berlin: Akademie-Verlag. [MR0818219](#)
- [19] LO, A.Y. (1984). On a class of Bayesian nonparametric estimates. I. Density estimates. *Ann. Statist.* **12** 351–357. [MR0733519](#)
- [20] MARTIN, R. and GHOSH, J.K. (2008). Stochastic approximation and Newton's estimate of a mixing distribution. *Statist. Sci.* **23** 365–382. [MR2483909](#)
- [21] MARTIN, R. and TOKDAR, S.T. (2009). Asymptotic properties of predictive recursion: Robustness and rate of convergence. *Electron. J. Stat.* **3** 1455–1472. [MR2578833](#)
- [22] MEYN, S.P. and TWEEDIE, R.L. (1993). *Markov Chains and Stochastic Stability. Communications and Control Engineering Series*. London: Springer London Ltd. [MR1287609](#)
- [23] NEWTON, M.A. (2002). On a nonparametric recursive estimator of the mixing distribution. *Sankhyā Ser. A* **64** 306–322. [MR1981761](#)

- [24] NEWTON, M.A., QUINTANA, F.A. and ZHANG, Y. (1998). Nonparametric Bayes methods using predictive updating. In *Practical Nonparametric and Semiparametric Bayesian Statistics* (D. DEY, P. MÜLLER, and D. SINHA, eds.). *Lecture Notes in Statist.* **133** 45–61. New York: Springer. [MR1630075](#)
- [25] NEWTON, M.A. and ZHANG, Y. (1999). A recursive algorithm for nonparametric analysis with missing data. *Biometrika* **86** 15–26. [MR1688068](#)
- [26] PITMAN, J. (2006). *Combinatorial Stochastic Processes. Lecture Notes in Math.* **1875**. Berlin: Springer. [MR2245368](#)
- [27] PROPP, J.G. and WILSON, D.B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* **9** 223–252. [MR1611693](#)
- [28] REGAZZINI, E., GUGLIELMI, A. and DI NUNNO, G. (2002). Theory and numerical analysis for exact distributions of functionals of a Dirichlet process. *Ann. Statist.* **30** 1376–1411. [MR1936323](#)
- [29] ROBERTS, G.O. and TWEEDIE, R.L. (1999). Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Process. Appl.* **80** 211–229. Corrigendum. *Stochastic Process. Appl.* **91** 337–338. [MR1682243](#)
- [30] ROBERTS, G.O. and TWEEDIE, R.L. (2000). Rates of convergence of stochastically monotone and continuous time Markov models. *J. Appl. Probab.* **37** 359–373. [MR1780996](#)
- [31] RUGGIERO, M. and WALKER, S.G. (2009). Bayesian nonparametric construction of the Fleming-Viot process with fertility selection. *Statist. Sinica* **19** 707–720. [MR2514183](#)
- [32] SCHERVISH, M.J. (1995). *Theory of Statistics. Springer Series in Statistics*. New York: Springer. [MR1354146](#)
- [33] SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4** 639–650. [MR1309433](#)
- [34] TOKDAR, S.T., MARTIN, R. and GHOSH, J.K. (2009). Consistency of a recursive estimate of mixing distributions. *Ann. Statist.* **37** 2502–2522. [MR2543700](#)
- [35] TWEEDIE, R.L. (1975). Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stochastic Process. Appl.* **3** 385–403. [MR0436324](#)
- [36] TWEEDIE, R.L. (1976). Criteria for classifying general Markov chains. *Adv. in Appl. Probab.* **8** 737–771. [MR0451409](#)
- [37] VERVAAT, W. (1979). On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Adv. in Appl. Probab.* **11** 750–783. [MR0544194](#)
- [38] WALKER, S.G., HATJISPYROS, S.J. and NICOLERIS, T. (2007). A Fleming-Viot process and Bayesian nonparametrics. *Ann. Appl. Probab.* **17** 67–80. [MR2292580](#)
- [39] YAMATO, H. (1984). Characteristic functions of means of distributions chosen from a Dirichlet process. *Ann. Probab.* **12** 262–267. [MR0723745](#)

Received February 2010 and revised October 2010